Changes in protein levels as markers of severe disease: an investigation of severe malaria

Thesis

UNRESTRICTED

Changes in Protein Levels as Markers of Severe Disease: An Investigation of Severe Malaria

EVELYN NUNGARI GITAU

A thesis submitted to the Open University for the degree of

DOCTOR OF PHILOSOPHY

LIFE SCIENCES DISCIPLINE

SPONSORING ESTABLISHMENT

Kenya Medical Research Institute (KEMRI)/Wellcome Trust Research Programme, Kilifi Kenya

COLLABORATING ESTABLISHMENT

Department of Biochemical and Molecular Parasitology, Liverpool School of Tropical Medicine, Liverpool UK

October 2007

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Compounds directly involved in the pathogenesis of cerebral malaria (CM) remain unclear due to lack of robust methods of identifying and quantifying proteins expressed in low abundance. New developments in proteomics have now made it possible to identify low abundant proteins and provided new tools for studying host-parasite interactions. With these new tools, it may be possible to identify proteomic signatures for patients with various complications associated with severe malaria.

A global proteomic strategy was used to identify differentially expressed proteins in archived plasma and CSF drawn from children diagnosed with cerebral malaria (CM) compared to those with acute bacterial meningitis (ABM) and slide negative encephalopathy (EN). Samples were first separated using two-dimensional gel electrophoresis (2-DE) or two-dimensional liquid chromatography (2D-LC) and analysed using mass spectrometry. The data collected was analyzed using various bio-informatics tools. Finally, a CM mass profile was created using MALDI-ToF mass spectrometry.

Averages of about 150 spots per gel were resolved from CSF from CM and EN patients and 80 spots from ABM patients. In the gels from the CM and EN groups, 45 human proteins were found whilst 20 human proteins were unique to ABM compared to CM. For CSF, a total of 202 human proteins were identified using the 2D-LC system. Of these 13 were unique to CM, 124 to ABM and 32 to EN. 6 proteins were found in both CM and ABM and 18 were found in EN and ABM. 9 proteins were common to all 3 disease groups. A total of 66 *P. falciparum* proteins were identified but of these 48 were hypothetical proteins. Of the non-hypothetical proteins, 2 were found in both CM and ABM and the rest were found only in ABM.

Results show that proteomics can be used to create protein profiles of different disease groups. Majority of the human proteins identified by 2-DE were found to be high abundant proteins found in CSF and plasma. The use of 2D-LC enabled the identification of more low abundant proteins but some of the *P. falciparum* proteins identified by 2-DE were not seen in the 2D-LC method. Majority of the human proteins found were acute phase response plasma proteins including common circulating proteins such as albumin and apolipoproteins, blood transporters and binding proteins, protease inhibitors, enzymes, cytokines and hormones, and channel and receptor -derived proteins. There seems to be a correlation between the number of proteins found in the CSF and the level of blood brain barrier break down.

Acknowledgements

I am highly indebted to several people who have immensely contributed to the successful completion of this thesis. First, I wish to sincerely thank my supervisors, Professors Stephen A. Ward (Liverpool School of Tropical Medicine, Liverpool UK), Charles R.J.C. Newton and Gilbert O. Kokwaro a (both of KEMRI/Wellcome Trust Programme, Kilifi, Kenya), for their excellent supervision, intellectual guidance and helpful discussions, and constant encouragement throughout my study period.

I'm also grateful to the entire Molecular & Biochemical Parasitology Research Group, Liverpool School of Tropical Medicine, Liverpool, UK, both for the technical and moral support during my stay in Liverpool. I would especially like to extend my thanks to Mrs Alison Ardrey and Samantha Ashleigh, Drs Yang Wu and Gian Carlo Biagini. I would also like to give a special thanks to Dr Nabeel Hilal for helping with the techniques and Mrs Jill Davies for help with the animal work. A special mention goes to Dr Isabella Ochola for all her support while in Liverpool

I am very grateful to Drs Alexis Nzila and Simon Ndirangu both of the Pharmacology department KEMRI/Wellcome Trust Programme, Kilifi for their unfailing support throughout the study period

I would like to express my gratitude to all the field workers and clinicians KEMRI/Wellcome Trust Programme, Kilifi who collected all the samples used in the study.

I acknowledge, with great pleasure, the assistance given to me by the KEMRI/Wellcome Trust Programme, both in Kilifi and Nairobi with a special mention to Mrs Pauline Lowe, Mrs Veronica Wambugu and Nasim Gitonga for their support and encouragement. I also

thank the ICT data management team in Kilifi for their support, especially; Lazarus Mramba and Dr Gregory Fegan.

This project was financially supported by the Wellcome Trust (UK), KEMRI/Wellcome Trust Programme and WHO (TDR/MIM) through a grant to Professor Gilbert O. Kokwaro.

A final thanks to my family and friends for being there through thick and thin: My husband Henry and my dear son Njau for being patient when Mum was not always there, my parents Catherine and Joseph Gitau for getting me this far, and my siblings Njeri and Muriuki. May God Bless you all.

Dedication

To the strong women in my life-Cucu and Mum you taught me that it is in endurance that we find strength

Commit to the Lord whatever you do, and your plans

will surely succeed (Proverbs 16:3)

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

1D-LC	One-dimensional gel electrophoresis
2D DIGE	Two-dimensional difference gel electrophoresis
2-DE	Two-dimensional gel electrophoresis
2D-LC	Two-dimensional liquid chromatography
ABC	Ammonium Bicarbonate
ABM	Acute bacterial meningitis
CAN	Acetonitrile
APCI	Atmospheric Pressure Chemical Pressure Ionisation
BBB	Blood brain barrier
BLAST	Basic Logical Alignment Search Tool
CM	Cerebral Malaria
CNS	Central Nervous System
CSF	Cerebral Spinal Fluid
DART	Direct Analysis in Real Time
DTT	Diothiothreitol
EBI	European Bioinformatics Institute
EMBL	European Molecular Laboratory
EN	Encephalopathy
ESI	Electrospray Ionisation
FAB	Fast Atom Bombardment
GO	Gene Ontologies
HPLC	High Pressure Liquid Chromatography
IAA	Iodoacetamide
IEF	Isoelectric Focusing
IEX	Ion Exchange
IL-1	Interleukin-1
kD	kilo Dalton
LAPs	Low abundant Proteins
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass spectrometry

m/z	mass to charge
MALDI-ToF	Matrix Assisted Laser Detection and Ionisation-Time of Flight
Mr	relative molecular mass
MS	Mass Spectrometry
MW	Molecular weight
NCBI	National Centre for Biotechnology
NH_4HCO_3	Ammonium Bicarbonate
NMR	Nuclear Magnetic Resonance
NPRBC	non parasitised red blood cell
PIR	Protein Information Resource
PIRSF	Protein Information Resource Super Family
PMF	Peptide mass fingerprinting
PR	Prostate
PRBC	Parasitised red blood cell
RPLC	Reversed phase liquid chromatography
RP-LC	Reversed phase liquid chromatography
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate-poly-acrylamide gel electrophoresis
SELDI	Surface Enehanced Laser Detector and Ionisation
SSC	Scientific Steering Committee
SZ	Seizure
TEMED	Tetramethylethylenediamine
TFA	Triflouroacetic Acid
TNF	Tumour Necrosis Factor

1 Introduction and Literature Review

1.1 Introduction

Malaria still causes many childhood deaths in Africa, but the mechanisms leading to severe malaria are poorly understood. Understanding the moleular basis of the pathogenesis is critical to the development of new strategies for treatment of malaria. With developments in proteomics and the completion of the human and *Plasmodium falciparum* genome projects, new tools for studying host-parasite interactions have become available. With these new tools, it may be possible to identify new proteomic signatures for patients and to use this information to develop more sensitive diagnostic methods for malaria, novel methods for determining disease prognosis and new therapeutic interventions.

Studies described in this thesis use global proteomic strategies to identify differentially expressed proteins produced at various stages of severe malaria, and also attempt to define a protein mass profile for cerebral malaria (CM) - the most severe neurological complication of malaria.

1.2 The Global Burden of Malaria

Malaria has plagued humans throughout recorded history and is the most important parasitic disease worldwide, accounting for about 300-500 million clinical episodes and about a million child deaths per year of which 90% are children under the age of 5yrs (WHO, 2003, Korenromp, 2005). About 80% of the infections and practically all the severe forms of malaria and deaths are caused by *Plasmodium falciparum,* the most virulent plasmodium species in man (Korenromp, 2005). Recent estimates indicated that there were 515 (range 300-600) million episodes of clinical *P. falciparum* malaria in 2002, of which 365 million (70%) were in Africa (Snow et al., 2005).

Malaria remains a daunting health problem in tropical countries, but lack of data makes it even more difficult to ascertain the exact numbers of childhood cases and deaths. About 75% of the 550 million people in Africa live in areas of highly endemic stable transmission. Another 18% live in epidemic prone areas where malaria transmission is seasonal and unstable, with all age groups being vulnerable to infection and disease (WHO 1996). The sub-Saharan African region accounts for 70% of the disease burden, and this is mainly in children below the age of five years (Snow et al., 2005)(**Figure 1.1**).

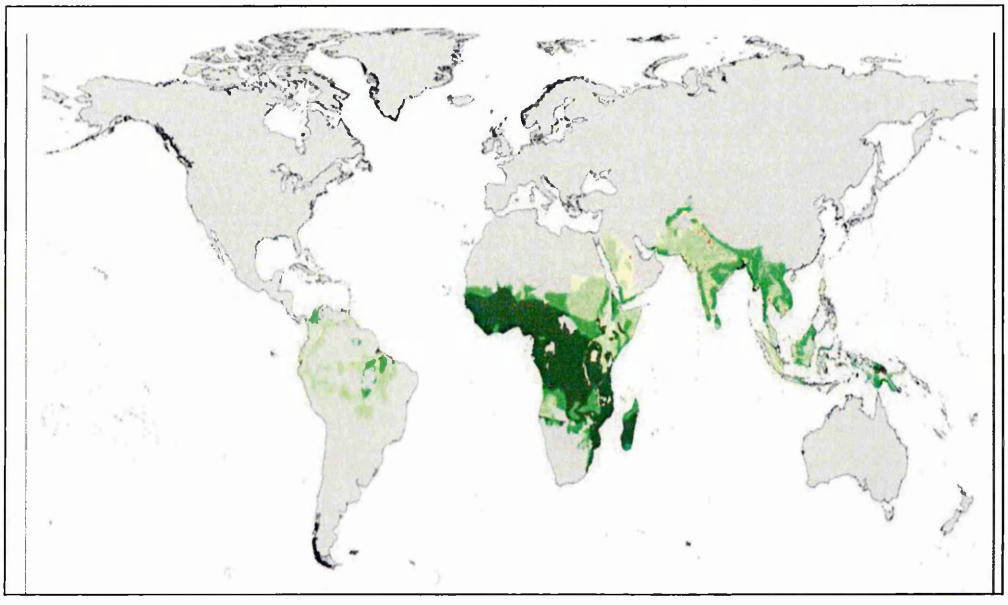

Figure 1.1 P. falciparum endemicity distribution within the global limits of risk. Endemicity classes: light green, hypoendemic (areas in which childhood infection prevalence is less than 10%); medium green, mesoendemic (areas with infection prevalence between 11% and 50%); dark green, hyperendemic and holoendemic (areas with an infection prevalence of 50% or more) Unclassified areas (yellow) represent only 6% of the global population at risk. Grey areas are a combined mask of areas outside of the transmission limits and areas of population density less than 1 person per km^2 Adapted from Snow *et. al.* (Snow et al., 2005).

P. falciparum malaria is the most important parasitic disease infecting the central nervous system (CNS) of humans worldwide. The parasite may infect humans at any time from conception to adulthood, but little is understood about the pathogenesis of the disease. Features of severe malaria include cerebral malaria (CM), severe anaemia, respiratory distress (acidosis), repeated convulsions, hypoglycaemia and hyperparasitaemia. The clinical symptoms and signs of falciparum malaria are non-specific and are characteristic

of a number of other febrile illnesses found in the malaria endemic areas. A key question in trying to understand the pathogenesis of severe malaria, and especially CM, is how the parasite developing within the erythrocytes, sequestered within the blood vessels, disrupts the function of tissues, particularly the CNS (Gitau and Newton, 2005). Studies aimed at answering this question may lead to development of new tool for treatment of malaria and management of associated complications.

Chemotherapy is the main control strategy for malaria; however, the development of drug resistance to most of the current drugs available is causing a crisis in the use and deployment of these drugs for prophylaxis and treatment of the disease (Cowman, 2001). Without effective control, clinical episodes of malaria will more than double over the next 20 years. The current problems with chemotherapy in malaria and the incomplete understanding of its pathophysiology justify the search for new tools for better diagnosis and treatment of malaria and associated complications.

1.3 The Malaria Parasite and its Life Cycle

Malaria is a disease caused by protozoan parasites belonging to the genus *Plasmodium*. Four species of *Plasmodium* cause disease in humans: *P. malariae* -Laveran, 1881, *P. vivax*-Grassi and Feletti, 1890, *P. falciparum* -Welch, 1897 and *P. ovale* - Stephens, 1922 (Bruce-Chwatt, 1988) but *P. falciparum* causes most problems as a result of its prevalence, virulence, pathophysiological consequences and drug resistance patterns. The *Plasmodium* life cycle takes place in two phases- a sexual replication stage (sporogony) within the mosquito and an asexual stage (schizogony) within the humans. In the first phase, the female mosquito ingests male and female gametocytes (micro- and macro-gametocytes, respectively) while taking a blood meal from an infected human. These undergo maturation within the mosquito, and then come together during fertilisation in the gut of the mosquito to form a globular zygote. The zygote, which is initially immotile, becomes elongated and motile to form an oökinete. The oökinete invades the midgut wall

of the mosquito to develop into a static oöcyst. After maturation, the oöcyst ruptures to release motile sporozoites which migrate to the mosquito salivary glands and are injected into humans during the next blood meal (Garnham, 1988a, Garnham, 1988b).

In the human host, sporozoites which survive the body's defences, invade the liver cells (exo-erythrocytic stage), which culminates in the rupture of the resultant schizont releasing thousands of merozoites into the blood stream. In *P. vivax* and *P. ovale,* some sporozoites remain in liver cells to form hypnozoites, which often lead to relapsing malaria. Circulating merozoites infect red blood cells (erythrocytic stage), undergo development into a schizont, which raptures releasing the merozoites into the plasma to infect more red cells. Some merozoites differentiate into gametocytes, which are ingested by mosquitoes during a blood meal; and undergo a cycle of maturation within the mosquito gut to form sporozoites, which infect hosts in the next blood meal (**Figure 1.2**).

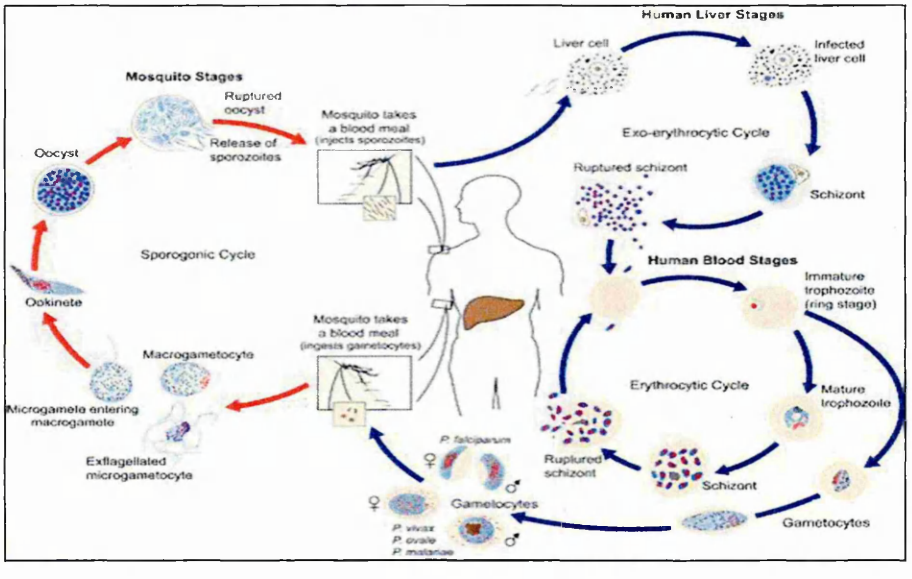

Figure 1.2 Life cycle of *Plasmodium* in man and in the mosquito (http://www.cdc.gov/malaria/biology/life_cycle.htm, accessed 27/07/07).

1.4 Pathogenesis of Severe Malaria

Understanding the pathogenesis of severe malaria is critical to improving its management. However, the pathogenesis of severe malaria is complex and remains poorly understood. The clinical symptoms and signs of falciparum malaria are non-specific and are characteristic of a number of other febrile illnesses found in the malaria endemic areas. A key question in the pathogenesis of severe malaria, and especially CM, is how the parasite developing within the erythrocytes, sequestered within the blood vessels, disrupts the function of tissues, particularly the (CNS), suggests that there may be factors associated with the presence of parasites within the vascular system that affect CNS function.

A study from Kilifi, Kenya suggests that *P. falciparum* isolates that bind to multiple receptors are involved in the causation of severe malaria and several receptor-ligand interactions work synergistically in bringing about severe disease (Heddini et al., 2001). Cell to cell interactions between the parasite and host involving adherence/invasion appear generally but not exclusively to correlate with severe disease (Pasvol, 2001). These cell interactions, including but not limited to rosetting, sequestration and the induction of pro-inflammatory cytokines by parasitised red blood cells (PRBC) are all associated with production of specific proteins and other compounds whose identification and quantification would help in understanding the pathogenesis of severe malaria. These compounds could be the result of parasite metabolism or host-specific changes as a result of parasite invasion. For example, in a pilot study using Nuclear Magnetic Resonance (NMR) spectroscopy on cerebral spinal fluid (CSF) from children with cerebral malaria, the spectra were different from patients with stroke or multiple sclerosis, and suggested the presence of unidentified compounds that could have arisen from the malaria parasite metabolic activity or from host-specific changes (**Figure 1.3**). The significance of these findings remains unknown, and are explored in this thesis.

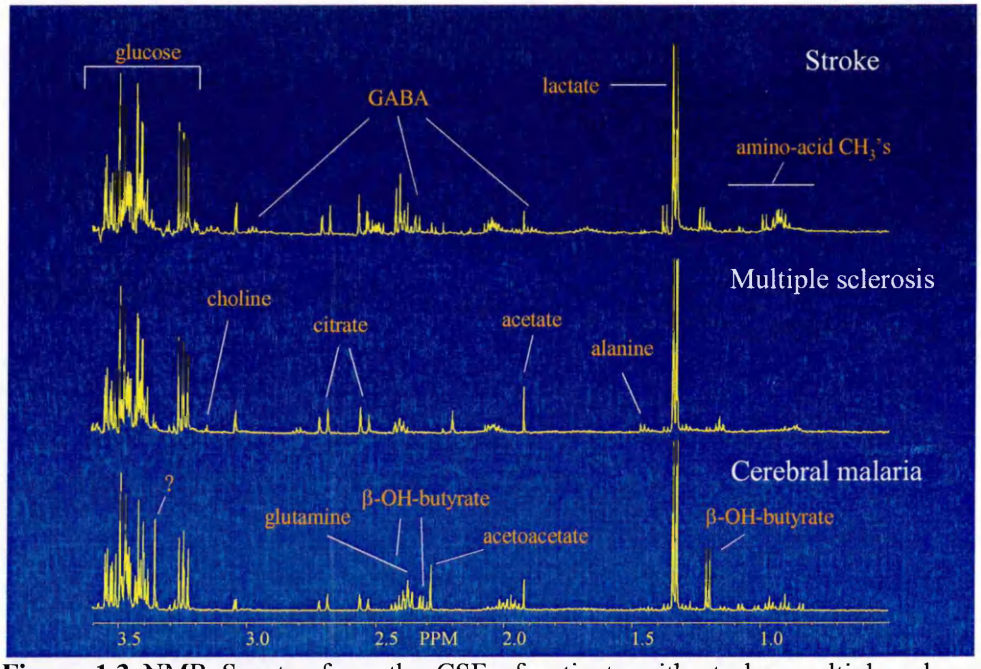

Figure 1.3 NMR Spectra from the CSF of patients with stroke, multiple sclerosis and cerebral malaria. The peak (?) one the spectra could not be assigned to a known compound (From: Parkes *et al.*, Unpublished data).

1.4.1 Neurological Complications of P. falciparum Malaria

Understanding the neurological complications of severe malaria may hold the key to the development of improved ways of management of malaria and associated complications, especially in children. CM is the most severe neurological syndrome of falciparum malaria with features of a diffuse encephalopathy which besides coma, also has brainstem signs and occasionally focal neurological signs (Newton and Warrell, 1998). There is a difference in the clinical complications between children living in endemic areas and non-immune adults.

CM in Adults

In adults, severe falciparum malaria is only seen in patients who have never acquired malaria or have lost their immunity (Newton and Warrell, 1998). The neurological manifestations of falciparum malaria in adults include seizures, agitation, psychosis, and impaired consciousness to coma. A number of post-malarial neurological syndromes (PMNS) have been described. Cerebellar ataxia following falciparum malaria has been

described in Sri Lanka (Senanayake and de Silva, 1994) but the pathogenesis is unclear. In Vietnam PMNS have been described where patients develop psychosis, seizures and tremors following symptomatic malaria (Nguyen et al., 1996)

CM in Children

In African children growing up in malaria endemic areas, severe *P. falciparum* malaria usually manifests as seizures, impaired consciousness or metabolic acidosis presenting as respiratory distress or severe anaemia (Marsh et al., 1995). CM is rarely encountered after the age of 10 years in people exposed to *P. falciparum* since birth. African children have a higher incidence of seizures than non–immune adults with evidence of brain stem injury (Molyneux et al., 1989, Marsh et al., 1995). Mortality ranges from 15-30% and 11% of children have neurological deficits on discharge (Newton and Krishna, 1998). This is likely to be an under estimate since recent studies suggest that up to a quarter have persistent deficits in language and other cognitive functions. The other neurological manifestations are usually transient but may also have sequelae. The pathogenesis of severe childhood malaria and mechanisms underlying impaired microvasculature circulation are briefly discussed below.

1.4.2 The Role of the Blood Brain Barrier (BBB) in Cerebral Malaria[1]

Various theories have been put forward to explain the pathogenesis of CM. In 1948 Maegraith and colleagues (Maegraith and Fletcher, 1972) proposed that the breakdown of the blood brain barrier (BBB) might cause CM. This 'permeability' hypothesis, based upon animal models of malaria (*vide infra*), suggested that a leaky BBB allowed compounds to enter the brain and thus cause neurological dysfunction. The 'mechanical theory' suggests that cytoadherence of PRBCs to endothelial cells (sequestration) leads to obstruction of the

[1] A review of the Blood Brain Barrier in *Falciparum* Malaria has been published in Tropical Medicine and International Health, see details in bibliography under Gitau and Newton (2005)

microcirculation and ischaemic anoxia. *In vitro* studies have shown that this is mediated by the interaction of a parasite protein *P. falciparum* membrane protein-1 (PfEMP-1) expressed on the surface of the infected erythrocytes (MacPherson et al., 1985, Warrell, 1987, Turner, 1997, Pongponratn et al., 1991). This theory has been disputed mainly because in some CM patients there has been a lack of sequestration and the fact that most patients recover without evidence of ischemic damage (Newton and Krishna, 1998).

It has also been hypothesised that sequestration is associated with activation of cerebral endothelial cells, as demonstrated by the up-regulation of cell surface antigens (Turner et al., 1994) and that PRBCs binding to receptors on cerebral endothelial cells in culture causes changes in the integrity of the BBB (Brown et al., 1999). The pathogenesis also features a systemic response characterised by increased levels of circulating cytokines (Grau et al., 1989). Receptor mediated signalling across the BBB then influences parenchymal cells to release local neuroactive mediators such as nitric oxide (NO), which is released by leukocytes and endothelial cells and contributes to coma in CM (Clark et al., 1993)

1.4.3 Rosetting and Aggregation

Rosetting is the adherence of non-parasitised red blood cells (NPRBCs) to PRBCs. Aggregation on the other hand is the clumping together of PRBCs in *ex vivo* cultures. An association between aggregation and severe disease has been reported (Pain et al., 2001). Both these phenomena may play a role in the pathogenesis of CM. Rosetting is pH and heparin sensitive and is dependant upon the presence of divalent cations (Carlson et al., 1990a, Carlson et al., 1990b) and the presence of ligands (PfEMP1, complement-receptor 1 (CR1) coded for by *var* genes (Rowe et al., 1997). Rosettes can be disrupted by antibodies to *P. falciparum*, glycosaminoglycans and sulfated glycoconjugates. Increased rosette formation was found in Gambian children with CM (Treutiger et al., 1992) with a corresponding lack of anti-rosetting antibodies. In Kenyan children with malaria, rosetting

was correlated with severity of disease and was influenced by ABO blood group type (Rowe et al., 1997). Whether rosetting and aggregation plays a direct role in the pathogenesis of severe malaria or is a marker for some other causal factor and the association with specific parasite phenotype remains unclear (Rowe et al., 1995, Roberts et al., 2000).

1.4.4 Sequestration

Mechanical obstruction of cerebral blood vessels as a cause of CM was first proposed by Lavaran in 1893. It has remained a favourable explanation but the mechanism is not well understood to date. Initially the obstruction was thought to be due to thrombi formation (Dudgeon and Clark, 1917), although the agglutinates in the vessels lack fibrin and platelets. The other view is that the cytoadherance of the PRBCs to the endothelium of cerebral capillaries causes obstruction (sequestration). Sequestration results from adhesive interactions between parasite derived proteins expressed on the surface of pRBCs and a number of host molecules on the surface of endothelial cells, placental cells and uninfected RBCs (Beeson and Brown, 2002). One of the best characterised ligands is *P. falciparum* erythrocyte membrane protein (PfEMP)-1, which is exported by the parasite onto the RBC membrane at adhesive loci called knobs (Adams et al., 2002).

1.1.1 Red cell deformability

This is a phenomenon where PRBCs become less flexible and are unable to pass through the microvasculature with diameters < 7.5 μm, the diameter of normal red blood cells. This has been reported in *P.falciparum* PRBC (Nash et al., 1989). This deformability could contribute to appreciable microcirculatory obstruction and grossly abnormal cells might cause delay at capillary entrances. Moreover, in vessels lined with adhered cells, the flow in the central region would also be impaired by the presence on any poorly deformed cells.

1.4.5 Cytokine Induction

(Clark et al., 1991) have suggested that host over production of pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin (IL)-1 are capable of inducing cerebral symptoms in patients with malaria. Induction of these inflammatory cytokines could be triggered by the PRBC or by parasite derived material (Newton and Krishna, 1998)

1.5 Murine Model for Cerebral Malaria

Some of the biochemical changes that occur in severe malaria (type, amount and location of production of protein) may vary with time. As such, it is difficult to fully characterise the production of such compounds from samples obtained from patients during routine clinical examination. An animal model of malaria may be more useful in the initial efforts to characterise the production of such proteins, even though they may not be the same compounds as found in the human form of malaria. One such animal model that is used to study certain aspects of severe malaria is the murine model. Cerebral malaria is induced in susceptible strains of mice following infection with *Plasmodium berghei* ANKA[2] (PbA) strain. This murine model of cerebral malaria has been used in the past to throw light on the pathogenesis of the human condition (Jennings et al., 1997, Neill et al., 1993, Reed et al., 1997, Rest, 1982, Thumwood et al., 1988). CBA/T6 inbred strain of mice infected with *P. berghei* ANKA strain (PbA-CBA) develop severe neurological symptoms and die between 6-9 days post infection (Jennings et al., 1997, Neill et al., 1993, Reed et al., 1997, Rest, 1982, Thumwood et al., 1988) whereas DBA/2J mice infected with the same (PbA-DBA) develop a non-fatal cerebral malaria (Neill et al., 1993). When infected with *P. berghei* K173 strain (Pb), both CBA and DBA mice develop a non-cerebral form of

---

[2] This name consists of a contraction of ANtwerpen, where the Prince Leopold Institute is, and KAsapa, the province of Katanga where the finding was made.

malaria and die 15-22 days after infection with the parasite, from profound anaemia, parasitaemia and monocytosis (Neill et al., 1993).

The argument that *P. berghei* ANKA CM is not a good model of *P. falciparum* CM because it is characterised by mononuclear cell inflammation rather than sequestration of parasitised erythrocytes has been challenged on many grounds. Although the experimental murine CM cannot exactly reproduce the brain pathology found in humans, there are some similarities between the pathologies seen in several models, and in particular the PbA model, and the human pathologies (Chang et al., 2001). Several observations from the study of experimental CM in mice have been extended and confirmed in human disease. For example, PRBC sequestration also occurs in murine CM although in a less prominent manner than in humans (Hearn et al., 2000). Other similarities include brain haemorrhage, plugging of microvessels, some necrosis of microvessels and upregulation of the production of pro-inflammatory cytokines especially TNF (Combes et al., 2006, Grau et al., 1987, Grau et al., 1988, Grau et al., 1989). Another study has demonstrated that mononuclear inflammation in murine models is associated with protection from lethal CM and may not be a part of CM's pathogenesis (Jennings et al. 1998). If such changes occurred in humans, it would be impossible to demonstrate because biopsies are not made during recovery from CM (Jennings et al. 1998). It appears that both the human and the *murine* CM involve more complex processes. The experimental model allows one to follow protein expression throughout the entire course of the disease.

In this thesis recent advances in proteomics have been applied to such animal models to gain further insight into the pathogenesis of severe malaria. The major difference between the proteomic techniques described in the thesis and the non-proteomic techniques currently in use, is that the proteomic approach allows identification and quantification of proteins expressed in small amounts even at times during the course of disease when no major pathological changes would have occurred.

1.6 Proteomics and Disease: Opportunities for Development of New Diagnositic Tools and Management Strategies for Malaria

Discussions in the preceding sections have shown that severe malaria may be associated with production of compounds including low abundant proteins (LAPS) from the parasite or host. Identification of such compounds may provide opportunities for refinement of current methods for diagnosis and management of malarial. One particular recent development that can be exploited in this regard is proteomics. Proteomics is the large-scale study of the structure and functions of proteins, and includes the rapidly evolving field of disease proteomics. It has been applied to the field of infectious diseases (Nally et al., 2005, Walduck et al., 2004) and this thesis describes application of recently developed proteomic techniques to gain further insight into the pathogenesis of severe malaria. The proteomic strategy and the technological basis of proteomics are briefly described in this section and are well described in chapter 2.

1.6.1 Proteomic Strategies for Detection of Disease-Specific Proteins

Proteome analysis is a direct measurement of proteins in terms of their presence and relative abundance, and is regarded as a technology independent from, but complementary to genome analysis. Proteins contain several dimensions that collectively indicate the actual rather than the potential functional state as indicated in mRNA analysis (Gygi et al., 2000b, Humphery-Smith et al., 1997, Haynes and Yates, 2000) and therefore genomic information alone cannot be extrapolated to represent the functional status of a biological system. The overall aim of a proteomic study is characterisation of the complex network of cell regulation by proteins.

There have been rapid advances in the field of molecular medicine, which may provide a molecular fingerprint of an individual patient and disease phenotype (Rohlff, 2000). Protein-targeting interactions play a central role in most biological processes, and their detection and analysis can provide important information on specificity, affinity and

structure-function relationships (Gromov et al. 2002). Proteome analysis is required to determine which proteins have been conditionally expressed, how strongly and whether any post translational modifications are affected. Briefly, proteomic analysis involves the following critical steps

Protein separation using two dimensional polyacrcylamide gelelectophoresis(2-DE)

A commonly used method to separate proteins according to their size and charge is 2D-E. First, proteins from biological sources are separated on a thin gel according to their isoelectric charge (first dimension). In the second dimension, the proteins are separated according to their size.

Protein Separation using liquid chromatography

Recent developments in liquid chromatography (LC) linked to mass spectrometry have further refined the separation of proteins from complex biological matrix. LC has been used in the past to separate proteins digested on gels but recent advances have enabled better detection of LAPs making it a viable alternative to 2-DE. Another attractive feature of LC is the broad selection of stationary and mobile phases enabling a better separation of highly acidic and basic proteins as well as hydrophobic proteins which are hard to detect on 2-DE (Gygi and Aebersold, 2000).

Mass spectrometry

The application of mass spectrometry to proteins has been made possible by the development of methods for ionising large biomolecules without destroying them. These include electrospray ionisation and matrix-assisted laser desorption ionization (MALDI). For mass spectrometry, a protein is digested with an enzyme (trypsin) and the resulting fragments energised by laser (e.g. MALDI). The charged peptide fragments are then separated according to their mass-to-charge ratio by various methods, e.g. "time of flight (ToF) before impinging on a detector which measures the signal intensity of each mass.

Bioinformatics for protein databases searching and protein cataloguing

The identification of proteins from peptide sequence derived from the mass spectrum has been facilitated by developments of databases for proteomics which has led to enormous quantity and variety of data being produced in the growing number of molecular databases. To fully these valuable data, advanced bioinformatics infrastructures have been developed for biological knowledge management. One major challenge lies in the volume, complexity, and dynamic nature of data being collected and maintained in heterogeneous and distributed sources. To facilitate scientific discovery, information scattered in disparate sources needs to be integrated into a cohesive framework. With data integration, interesting relationships among protein family, structure, and function can be readily revealed, providing for plausible function and pathway identification.

Applications of proteomics

Identification of the proteins expressed in severe disease can be used in effectively confirming suspected protein interactions as well as in identifying new binding partners.

Identification of biomarkers using a proteomic strategy can help gain a mechanistic understanding of the differences in clinical responses to therapeutic treatment. The biomarkers could also help in evaluating toxicity, drug efficacy or exposure to xenobiotics in man or animals (Kennedy 2001). Biomarker measurements can also help explain empirical results of clinical trials by relating the effects of interventions on molecular and cellular pathways to clinical responses (Atkinson et al., 2001). Proteomic patterns have been used in the past to predict the risk of development of ovarian cancer in high risk groups and in the general population (Petricoin et al., 2002a) In this study, proteomic spectra were generated by mass spectrometry. A preliminary "training" set of spectra derived from patients with ovarian cancer were analysed by an iterative searching algorithm that identified a proteomic pattern that completely discriminated cancer from non-cancer cases. The discovered pattern was then used to classify an independent set of

masked serum samples. The algorithm identified a cluster pattern that, in the training set completely segregated cancer from non-cancer including patients in stage I cancer (Petricoin et al., 2002c). As another example, it was recently reported that it is possible to identify patients with sleeping sickness based on a unique proteomic signature (Papadopoulos et al 2004). These findings justify a prospective population based assessment of proteomic pattern technology as a screening tool for stages of this disease. Biomarker measurements can also help explain empirical results of clinical trials by relating the effects of interventions on molecular and cellular pathways to clinical response (Atkinson et al., 2001).

The Plasma Proteome

Plasma is the longest studied diagnostic proteome mainly because it is readily and safely available in large amounts (70 mg/ml) (Putnam F., 1975). The human plasma proteome is likely to contain most, if not all, human proteins, as well as proteins derived from some infective organisms such as viruses, bacteria, and fungi. Many of the human proteins, introduced by low-level tissue leakage, ought to be present at very low concentrations while others, such as albumin, are present in very large amounts (Anderson et al., 2004). From a functional viewpoint proteins present in plasma can be grouped into various groups as listed in **Table 1.1.** Numerous post-translational modified forms of each protein are likely to be present, along with millions of distinct clonal immunoglobulin (Ig) sequences. This complexity and enormous dynamic range (**Figure 1.5**) make plasma the most difficult specimen to be dealt with by proteomics (Anderson and Anderson, 2002).

Table 1.1 Classification of Proteins found in Plasma as in described in Anderson and Anderson (Anderson and Anderson, 2002)

Proteins Secreted by Solid Tissues and That Act in Plasma—*The classical plasma proteins are largely secreted by the liver and intestines. A key aspect of plasma proteins is a native molecular mass larger than the kidney filtration cutoff (45 kDa) and thus an extended residence time in plasma (albumin, which is just larger than the cutoff, has a lifetime of about 21 days).*
Immunoglobulins—*Although the antibodies typically function in plasma, they represent a unique class of proteins because of their complexity: there are thought to be on the order of 10 million different sequences of antibodies in circulation in a normal adult.*
"Long Distance" Receptor Ligands—*The classical peptide and protein hormones are included in this group. These proteins come in a range of sizes, which may indicate a range of time scales for their control actions (i.e. rapid adjustment with small hormones such as insulin and slower adjustments with larger hormones such as erythropoietin).*
"Local" Receptor Ligands—*These include cytokines and other short distance mediators of cellular responses. In general these proteins have native molecular weights under the kidney filtration cutoff (and hence relatively short residence times in plasma) and appear to be designed to mediate local interactions between cells followed by dilution into plasma at ineffective levels. High plasma levels may cause deleterious effects remote from the site of synthesis, e.g. sepsis.*
Temporary Passengers—*These include non-hormone proteins that traverse the plasma compartment temporarily on their way to their site of primary function, e.g. lysosomal proteins that are secreted and then taken up via a receptor for sequestration in the lysosomes.*
Tissue Leakage Products—*These are proteins that normally function within cells but can be released into plasma as a result of cell death or damage. These proteins include many of the most important diagnostic markers, e.g. cardiac troponins, creatine kinase, or myoglobin used in the diagnosis of myocardial infarction.*
Aberrant Secretions—*These proteins are released from tumours and other diseased tissues, presumably not as a result of a functional requirement of the organism. These include cancer markers, which may be normal, non-plasma-accessible proteins expressed, secreted, or released into plasma by tumor cells.*
Foreign Proteins—*These are proteins of infectious organisms or parasites that are released into, or exposed to, the circulation.*

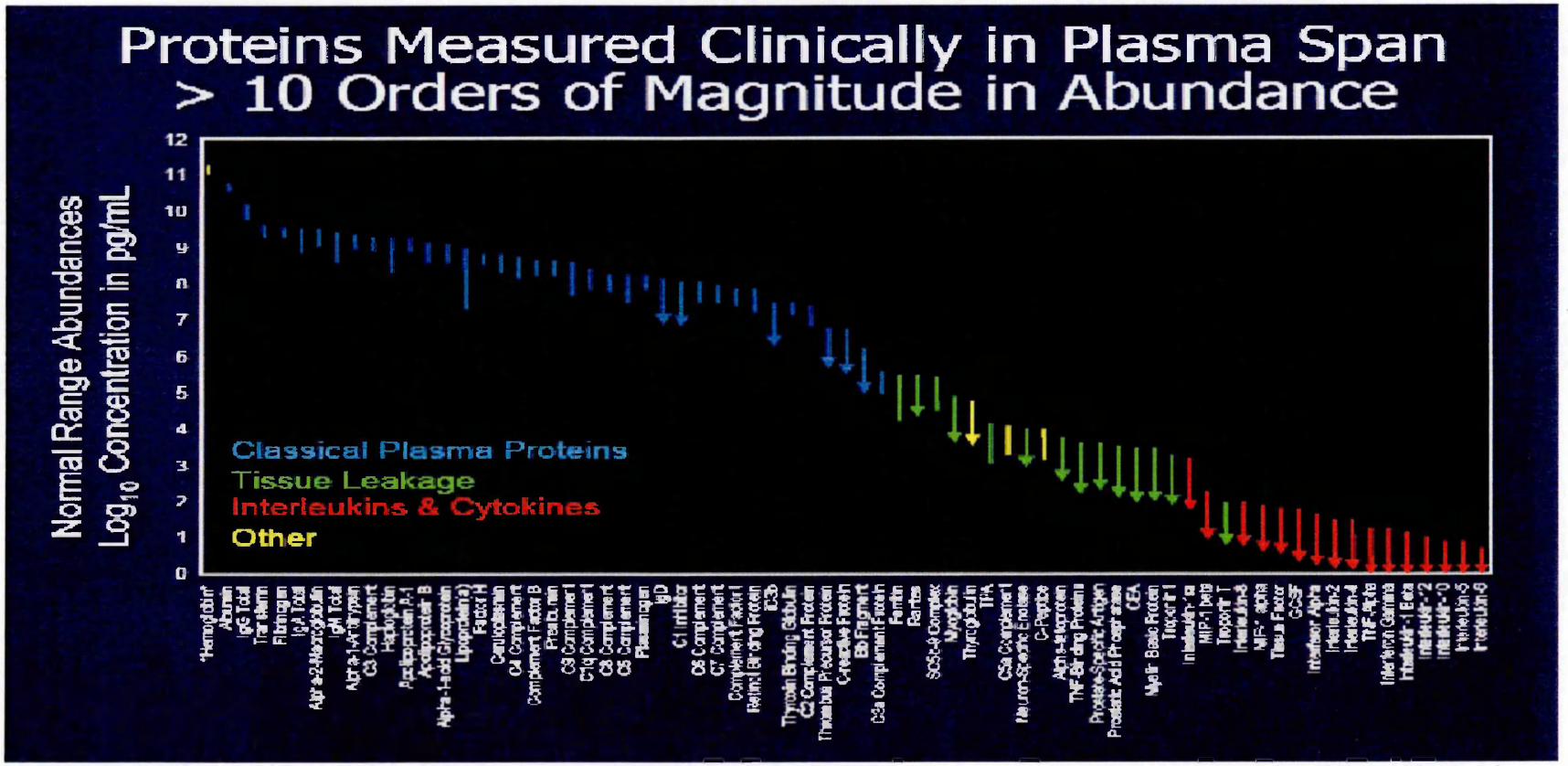

Figure 1.4 Reference intervals for 70 protein analytes in plasma. Abundance is plotted on a log scale spanning 12 orders of magnitude. Where only an upper limit is quoted, the *lower end* of the interval line shows an *arrowhead*. The classical plasma proteins are clustered to the *left* (high abundance), the tissue leakage markers (*e.g.* enzymes and troponins) are clustered in the *center*, and cytokines are clustered to the *right* (low abundance). Haemoglobin is included (*far left*) for comparison.). *TPA*, tissue plasminogen activator; *GCSF*, granulocyte colony-stimulating factor; *TNF*, tumor necrosis factor. Figure was adapted from the Plasma Proteome Institute (http://www.plasmaproteome.org/plasmaframes.htm)

Cerebrospinal fluid (CSF) is a clear fluid (colour changes usually indicate disease as shown on **Table 1.2)** that functions to cushion and protect the brain from changes in blood pressure and trauma. In addition the CSF transports neurosecreted, biosynthesised and metabolised cellular products (Romeo et al., 2005). CSF is not a filtrate of plasma. It is a secreted product of the highly vascular choroid plexus and the ependymal lining of the ventricles, which is reabsorbed by the arachnoid villa (Treseler, 1995, Yuan et al., 2002).

Table 1.2 CSF Supernatant Colours and Associated Conditions or Causes (Seehusen et al., 2003)

Color of CSF supernatant	*Conditions or causes*
Yellow	Blood breakdown products Hyperbilirubinemia CSF protein ≥150 mg per dL (1.5 g per L) >100,000 red blood cells per mm^3
Orange	Blood breakdown products High carotenoid ingestion
Pink	Blood breakdown products
Green	Hyperbilirubinemia Purulent CSF
Brown	Meningeal melanomatosis

Analysis of proteins in CSF is of great diagnostic importance because of its close proximity to the brain and its clinical availability (Davidsson et al. 2001). Diseases involving the CNS markedly affect the protein concentration and protein pattern of the CSF (Andersson et al. 1994; Andreasen et al. 1999). However, similar to plasma, the predominant proteins in CSF are isoforms of serum albumin, transferrin and immunoglobulins, which represent more than 70% of the total protein amount. Furthermore, an unwanted high dynamic range of protein abundance is found in CSF, making the detection of LAPs extremely challenging with the current analytical methods. An additional challenge with analyzing CSF is protein concentration. On average CSF contains 100 fold less protein than plasma, therefore, necessitating the need for larger sample amounts relative to plasma.

1.7 Scope of the Thesis

This thesis describes studies on the use of global proteomic strategies to determine the changes in protein levels in severe malaria. The hypothesis is that the intravascular parasites produce proteins that interfere with the function of tissues in which the parasites are sequestered. Some proteins may be produced in excessive amounts (e.g. excitotoxins) or others may be unique to the malaria infection.

1.1.2 Components of the thesis

The following are the broad components of the thesis:

Chapter 1 describes the basic biology of malaria transmission and the malaria endemicity and population with the focus on sub-Saharan Africa. It describes the pathogenesis of severe malaria with emphasis on the different hypothesis' for the mechanisms leading to the impaired neurological syndrome *P. falciparum* malaria. The chapter ends with a brief introduction to global proteomic strategies used in the studies described in this thesis.

Chapter 2 describes the materials and methods used in global proteomics. The chapter provides detailed background information to the different strategies employed in the thesis and also includes the experimental protocols used for each strategy employed.

Chapter 3 presents results of the global proteomic analysis of plasma from mice infected with *Plasmodium Berghei* ANKA strain using 2-DE and MALDI-ToF MS. The purpose of this analysis was mainly to validate the 2-DE method to be later used on archived human plasma and CSF samples of children diagnosed as having CM in Kilifi. Both chapters 3 and 4 end with a discussion of the advantages and disadvantages of the use of 2-DE on plasma and CSF and justifies the development of gel-free strategies in proteomics.

Chapter 4 gives a summary of the results of proteomic analysis of human (plasma and CSF) samples from African children with severe malaria.

Chapter 5 describes the use of a gel free proteomic strategy using multiple steps of LC prior to MS. It describes various strategies that have been employed previously and describes a modified 'shotgun' method of proteomics. The chapter ends with a detailed discussion of the advantages of this method over 2-DE but also discusses alternative modifications to the 'shotgun' method.

Chapter 6 describes the development of a method for mass profiling of CM using MALDI-ToF MS. The chapter reviews the various strategies of mass profiling currently employed and their pitfalls. It also demonstrates the difference and similarities between CM and other severe diseases commonly diagnosed in Kilifi including acute bacterial meningitis (ABM), and other encephalopathies. The chapter concludes by discussing the importance of disease profiling and its uses in prognosis, diagnosis as well as in defining treatment strategies.

Chapter 7 provides a summary of the thesis emphasising the major findings, overall importance of the findings in terms of strategies of prognosis and diagnosis and identifies some important areas that demand further investigation.

1.1.3 Objectives

The main objective the work described in the thesis was to employ proteomic strategies to define the changes in protein levels in severe malaria, with the specific objectives as follows.

1. Separate differentially expressed proteins in severe disease by use of 2-DE and LC.
2. Identify these proteins using MALDI-ToF and tandem mass spectrometry on an ion trap instrument and analyse using bio-informatics sources currently available to us.
3. Identify the proteins produced by the host/parasite in severe malaria.
4. Compare proteomes of CM, EN and ABM.

The work described in the thesis, therefore, used a variety of research methods to try and better understand the pathogenesis of severe malaria. The results will be used to establish a prospective study. Results can also be combined with other studies in search of immunogenic proteins for vaccine design.

2 Background and Experimental Methods in Proteomics

The main goal of the studies described in this thesis is to try and find proteomic signatures from plasma and CSF to define the changes in protein levels in severe malaria, which entails using techniques used to analyse proteins in complex mixtures. The strategies explored in the thesis include separation of the proteins using 2 dimensional gel electrophoresis and liquid chromatography mass spectrometry to analyse and bioinformatics tools to identify the proteins. These techniques as employed in proteomic analysis are described in detail in this chapter.

2.1 Two Dimensional Gel Electrophoresis

2.1.1 Background

Two dimensional gel electrophoresis (2-DE), is a tool widely used for analyzing complex mixtures from cells, tissues and other biological samples. This technique has been applied to study protein profiles of CSF and plasma to identify disease related alterations in protein expression in severe malaria disease. High-resolution two dimensional gel electrophoresis (O'Farrell, 1975) using immobilized pH gradients (IPG) (Gorg et al., 2000) was employed to resolve and array quantitative expression profiling of large sets of complex protein mixtures according to their isoelectric point (pI), and relative molecular mass (M_r). Subsequent analysis of the visualised protein expression patterns is performed using image analysis software, and protein spot identification is then achieved using mass spectrometry coupled with genomic database comparisons.

In 2-DE, proteins are separated according to their pI by isoelectric focusing (IEF) in the first dimension and according to their M_r by sodium dodecyl sulphate poly-acrylamide gel electrophoresis (SDS-PAGE) in the second dimension. Since these two parameters are unrelated, it is possible to obtain an almost uniform distribution of protein spots across a 2-

DE gel. The resolving capacity of the 2-DE gels is dependent on the separation length in both dimensions and is usually considered to be proportional to the total gel area available for separation. Each spot on the resulting two-dimensional array may correspond to a single protein species in the sample, but quite often may be a mixture of proteins. This method is characterized by a number of important features; (a) extremely high resolving power, (b) tolerance to crude protein mixtures, (c) tolerance to relatively high sample loads, (d) efficient fractionation of sample, and (e) separated proteins are conserved in the gel matrix for further analysis at any desired time (Gorg and Weiss, 1999).

In the studies described in this thesis, 2-DE using IPGs was used instead of the classic 2-DE, which uses carrier ampholytes (Gorg et al., 2000). The former has the advantage of higher resolution and higher loading capacity; cathodic drift is completely abolished improving reproducibility and there is uniform conductivity and buffering capacity which extends the basic pH limit for 2-DE (Bjellqvist et al., 1982). In brief, 2-DE using IPGs involves: (a) sample preparation, (b) IPG strip rehydration, (c) IEF (d) IPG strip equilibration (e) SDS-PAGE, (f) Detection and visualization of spots and (g) Protein spot identification. Details of the application of this technique in proteomic analysis are given below

2.1.2 Sample Preparation

Sample preparation is a vital factor for the overall performance of the 2-DE technique (Thongboonkerd et al., 2004, Shaw and Riederer, 2003). No single method of sample preparation can be applied due to the diverse nature of samples. However some general considerations apply to all samples. It is essential to minimise, protein disaggregation and removal of interfering components (such as nucleic acids, and lipids), and to minimise proteolytic degradation, artifactual oxidation, carbamylation, and conformational alteration of proteins (co-analytical modifications) that might lead to artifactual spots on 2-DE gel profiles (Rabilloud, 1996).

Soluble, liquid samples such as plasma and CSF can often be analysed by 2-DE with minimal pre-treatment. However in the case of CSF the relatively low protein concentration and high salt content can interfere with protein separation during IEF. Such samples can be desalted by dialysis or liquid chromatography prior to 2-DE (Gorg et al., 1997).

For solid tissue samples it is necessary to maximise solubilisation by breaking molecular interactions, including disulphide bridges, the main forces holding proteins together; non-covalent interactions like ionic bonds, hydrogen bonds, and hydrophobic interactions (between proteins and other compounds); and covalent bonds found mainly between proteins and some coenzymes. Urea is used as a denaturant in the first dimension of 2 DE. Thiourea is a much stronger denaturant than urea, but cannot be used alone as it is weakly soluble in water. However, it is more soluble in a concentrated solution of urea, so that a mixture of urea-thiourea exhibits increased solubilising power (Molloy et al., 1998, Rabilloud, 1998, Rabilloud et al., 1997). To ensure complete sample solubilisation of the hydrophobic residues exposed as a result of chaotropic denaturation, and prevention of aggregation through hydrophobic interactions, a non-ionic or zwitterionic detergent (surfactant) is always included in the sample solution. In recent years the sulphobetaine CHAPS, a zwitterionic detergent, has become the surfactant of choice and is generally used at between 2-5% in 8 M urea (Herbert, 1999) to increase the solubility of hydrophobic proteins. Traditionally, nonionic polyol mixtures such as the Triton X-100 and Nonidet P-40 have been used (O'Farrell, 1975, Klose and Spielmann, 1975), but are of less purity compared to CHAPS. New sulphobetaine surfactants such as N-decyl-N,Ndimethyl-3-ammonio-1-propane sulphonate (SB 3-10), when used in combination with thiourea-urea mixture, provide a wide range of powerful sample solutions for 2-DE. However, they suffer from poor solubility in high concentrations (> 5M) of urea (Rabilloud et al., 1990). SDS, a commonly used ionic detergent, is poorly compatible with IEF, though low

amounts may be used in the initial sample solubilisation prior to IEF, provided that high urea concentrations and non-ionic or zwitterionic detergents are present to ensure complete removal of the SDS from the proteins during IEF. The final concentration of SDS should be 0.25% or lower, and the ratio of the excess detergent to SDS should be at least 8:1 (Ames and Nikaido, 1976, Wilson et al., 1977).

Inclusion of reducing agents in the sample solution helps to break any disulphide bonds present and to maintain all proteins in their completely reduced state. Addition of an excess of a sulphydryl (free thiol containing) reductant such as dithiothreitol (DTT) at concentrations ranging from 20 to 100 mM is commonly used. Dithierythritol structurally akin to DTT is an alternative. Non-thiol reductant Trialkyl phosphines such as tributyl phosphine (TBP), have also been reported to increase the solubility of proteins during the IEF and to increase transfer of proteins to the to the second dimension (Herbert, 1999, Herbert et al., 1998). Addition of carrier ampholytes or IPG buffer (0.5-2% v/v) can be beneficial, enhancing protein solubility by minimising protein aggregation due to charge-to charge interactions and scavenging cyanate ions (Barrett et al., 2000). They do not disturb the IEF very much because they migrate to their pIs, where they become uncharged. In addition to chemical solubilisation techniques, mechanical procedures of cell disruption like trituration (cells lysed by shear forces resulting from forcing cell sample through a small orifice under high pressure) and sonication (in short bursts to avoid heating), facilitate disaggregation of protein molecules.

Protein degradation or modification by proteolysis must be prevented during sample preparation. This can be achieved by disrupting the sample using strong denaturants such as 8M Urea, 10% TCA or 2% SDS (Harrison and Black, 1982, Wu and Wang, 1984, Granzier and Wang, 1993) or by adding commercially available protease cocktail inhibitors. Additionally proteolysis can be inhibited by preparing samples at pH 9 and

above by using Tris base, sodium carbonate or basic carrier ampholytes (Berkelman and Lagarias, 1986).

Presence of nucleic acids, especially DNA may affect the separation of proteins by IEF. DNA complexes are dissociated during denaturing conditions as occurs in sample preparation. This inhibits protein entry and slows migration in the IPG gel, and furthermore DNA binds to proteins in the sample and causes artifactual migration and streaking (Rabilloud, 1996). Nucleic acids are efficiently broken down by sonication and may also be degraded by addition of a suitable pure (i.e. protease-free) DNase/RNase endonuclease mixture to the sample solubilisation solution. An alternative method is to utilise the ability of synthetic carrier ampholytes to form complexes with nucleic acids and then remove the complexes by ultracentrifugation.

2.1.3 First dimension isoelectric focusing (IEF)

Proteins are positively charged at pH values less than their pI and negatively charged at pH values greater than their pI. IEF separates proteins in a pH gradient until they reach a stationary position where their net charge is zero (pI). This is achieved by subjecting the proteins to an electric field; those with a positive net charge migrate to the cathode while those with a negative net charge migrate towards the anode, until they reach their pI.

Sample application may be achieved passively by either including it in the rehydration solution or by direct application to the rehydrated IPG strip using sample cups or sample wells. In-gel sample rehydration or rehydration-loading of the IPG is the preferred method of sample application, where the protein sample is included in the rehydration solution, using the entire IPG gel for sample application, as suggested by (Rabilloud et al., 1994a). This has been shown to eliminate the formation of precipitates at the application point often associated with loading using sample cups, because there is no discrete application point resulting in improved resolution throughout the pH range of the gel, and allows

precise control of protein amounts and sample volumes loaded into the IPG gels. The method is also technically simpler, avoiding problems of leakage that can occur when using sample cups. Regardless of where proteins start in the pH gradient, they migrate in the electric field to their corresponding isoelectric points (pIs).

IPG strips are rehydrated prior to IEF. The specific protein solubility requirements of the sample determine the choice of the optimal rehydration solution. Generally, a typical rehydration solution consists of a chaotropic agent for solubilising and denaturing proteins, unfolding them to expose internal ionisable amino acids. Addition of thiourea to urea, has been reported to improve membrane protein solubilisation (Rabilloud et al., 1997, Rabilloud, 1998, Molloy et al., 1998). Non-ionic and zwitterionic detergents like CHAPS, Triton X-100, or NP-40 usually in concentrations of 0.5 to 4% are also added to solubilise hydrophobic proteins and minimise protein aggregation. To cleave the disulphide bonds, a reducing agent like DTT or DTE is commonly added in concentrations of 20-100 mM. Use of non-thiol reductant tributyl phosphine has also been documented (Herbert et al., 1998). To enhance the sample solubility and produce uniform conductitvity across the pH gradient, a carrier ampholyte mixture or IPG buffer of desired pH range is added. Lastly, a tracking dye (bromophenol blue) is added to help visualise IEF progress at the beginning of focusing.

2.1.4 IPG strip equilibration

Prior to transfer of the IPG strip onto the SDS gel, the strips are treated with SDS along with other reagents. This denatures the proteins (that is, it unfolds them into long, straight molecules) and binds a number of SDS molecules roughly proportional to the protein's length and because a protein's length (when unfolded) is roughly proportional to its mass, this is equivalent to saying that it attaches a number of SDS molecules roughly proportional to the protein's mass. Since the SDS molecules are negatively charged, the result of this is that all of the proteins will have approximately the same mass-to-charge

ratio as each other. In addition, proteins will not migrate when they have no charge (a result of the isoelectric focusing step) therefore the coating of the protein in SDS (negatively charged) allows migration of the proteins in the second dimension. This equilibration step saturates the IPG with the SDS buffer system required for the second dimension separation. Special precautions are taken to minimise protein losses during this stage, while maximising transfer of proteins on to the second dimension, and these are reflected in the modified composition of the equilibration buffer (Gorg et al., 1995).

2.1.5 Second Dimension SDS-PAGE

SDS PAGE is an electrophoresis method for separating polypeptides according to their molecular weight (MW). SDS-PAGE consists of three steps: (a) second-dimension gel preparation, (b) transfer of the equilibrated IPG strip on to the SDS gel, and (c) electrophoresis. The technique requires that all the proteins in the mixture have the same net charge per gram, for the movement through the gel to be solely based on the molecular mass of the proteins. This is achieved by performing the separation in polyacrylamide gels containing SD). The charge modifier SDS is an anionic detergent (surfactant) that denatures proteins by binding to them. SDS inundates the intrinsic charge of proteins, thus disrupting hydrogen bonds and hydrophobic interactions and preventing protein aggregation, such that they all have the same charge density and free solution electrophoretic mobility. Addition of a reducing agent to cleave disulphide bonds, totally unfolds the protein molecules. When proteins are treated with both SDS and a reducing agent such as DTT, separations exclusively by molecular weight are possible. There is a log-linear relationship between the relative distance of migration of the SDS-polypeptide micelle and the molecular weight, for a certain molecular weight range depending on the polyacrylamide percentage used (Berkelman and Lagarias, 1986). The polyacrylamide gels are made by an aqueous free radical polymerization of the monomer acrylamide and the cross-linker *N,N'*-methylenebisacrylamide (Bis), using a redox initiator system such as

ammonium persulphate and *N,N,N′,N′*-tetramethylethylenediamine (TEMED) (Hames, 1994). Commonly SDS-PAGE in a Trischloride/ Tris-glycine buffer (Laemmli, 1970) is employed. A stacking gel is not needed in 2-DE protein separation, because the proteins are pre-separated by IEF and migrate from a gel into another gel. Vertical and horizontal flatbed systems can be used with similar results (Gorg et al., 1995).

2.1.6 Detection Techniques

Most detection methods used for SDS gels can be applied to second dimension gels. Features that characterise an ideal detection technique include; a wide linear dynamic range, high sensitivity, ability to be quantitative, compatibility with further analysis using mass spectrometry, fastness, non-toxicity, affordability (price) and non-dependence on living cells for labelling. None of the available techniques combines all these features. Autoradiography and fluorography are the most sensitive detection methods, although silver and Coomassie Brilliant blue staining are the most frequently used methods.

The most sensitive non-radioactive and widely used technique to detect protein spots on two-dimensional gels is silver staining, picking up protein amounts down to 0.2 ng. It is however much less stoichiometric, with the intensity being linear over 40-50 fold range in concentration from 0.04 ng/mm^2 to 2 ng /mm^2. There are two main types of silver stains employed for 2-DE gels: silver nitrate and silver diamidine. The latter shows better sensitivity for basic proteins, but is incompatible with tricine buffer in the gel, and silver mirror development on the gel surface is a common feature. In the silver nitrate method used in the studies described in this thesis, the silver is more weakly bound to the proteins (silver binds to the amino acid side chains, primarily the sulphydryl and carboxyl groups of proteins), and it can be modified for mass spectrometry compatibility (Shevchenko et al., 1996, Wilm et al., 1996). Silver staining is a complex, multi-step process, and many variables can influence the outcome. High purity reagents (including water used for

preparing the staining reagents) and precise timing are necessary for reproducible, high-quality results.

The Coomassie Brilliant blue staining, although 50-fold less sensitive than silver staining, is relatively simpler to use with a quantitative 20-fold linear range of 10-200ng (Brush et al., 1998). It is also more compatible with mass spectrometry. Colloidal Coomassie stain as described by (Neuhoff et al., 1988) is highly sensitive and is more commonly used. Negative staining with imidazole zinc is very sensitive, detecting protein amounts down to 0.2 ng. This method stains only the background, and not the proteins and thus provides a very good recovery yield for further analysis with mass spectrometry (Matsui et al., 1997). The down side of it is that it cannot be used for quantification.

Fluorescence staining methods are less sensitive than silver staining detecting protein amounts down to 2-8 ng. They however have broad linear dynamic ranges of about 100-fold , and they are compatible with subsequent mass spectrometry analysis with enhanced recovery of peptides from in-gel digests (Lopez et al., 2000). Currently available fluorescent dye stains include SYPRO® Ruby (most sensitive), SYPRO® Red and Orange. Unfortunately all these dyes are relatively expensive, and a fluorescence scanner or a charged couple device camera is required. Fluorescence labeling with Bimane or Cyanine dyes prior to isoelectric focusing gives similar sensitivities and dynamic ranges to fluorescence staining (Unlu et al., 1997, Tonge et al., 2001). Two-dimensional difference gel electrophoresis (2D DIGE) builds on this technique by adding a highly accurate quantitative dimension. 2D DIGE enables multiple protein samples to be separated on the same 2-DE gel. This is made possible by labelling of each sample using spectrally resolvable, size and charge-matched fluorescent dyes known as CyDye DIGE fluors (Patton, 2002). 2D DIGE involves use of a reference sample, known as an internal standard, which comprises equal amounts of all biological samples in the experiment. Including the internal standard on each gel in the experiment with the individual biological

samples means that the abundance of each protein spot on a gel can be measured relative (i.e. as a ratio) to its corresponding spot in the internal standard present on the same gel. Ettan DIGE is the system of technologies that has been optimized to fully benefit from the advantages provided by 2D DIGE (Marouga et al., 2005).

2.1.7 Image Analysis

2-DE research utilises software-based image analysis tools. These tools primarily analyse bio-markers by quantifying individual proteins, and showing the separation between one or more protein "spots" on a scanned image of a 2-DE product. Additionally, these tools enable one to match identical spots in serial gels and normalise the gels to compensate for non-expression related variations in protein spot intensity. Software packages include Delta2D® (Decodon), PDQuest® (Biorad) and Progenesis® software (Nonlinear Dynamics, Newcastle upon Tyne, UK) among others. In our studies, we have utilised both PDQuest® software version 6.2.1 and Progenesis PG 220® V2006 software (Nonlinear Dynamics, Newcastle upon Tyne, UK).

2.1.8 Experimental Protocols

2.1.8.1 Protein estimation

Protein determination was performed using the Bio-Rad Protein Assay, a dye-binding assay in which a differential colour change of a dye occurs in response to various concentrations of protein (Bradford, 1976). The absorbance maximum for an acidic solution of Coomassie® Brilliant Blue G-250 dye shifts from 465 nm to 595 nm when binding to protein occurs (Reisner et al., 1975). The Coomassie blue dye binds to primarily basic and aromatic amino acid residues, especially arginine (Compton and Jones, 1985). Spector (Spector, 1978) found that the extinction coefficient of a dye-albumin complex solution was constant over a 10-fold concentration range. Thus, Beer's law may

be applied for accurate quantitation of protein by selecting an appropriate ratio of dye volume to sample concentration. The Bio-Rad™ micro titre assay was used as follows:

a. The dye reagent was prepared by diluting 1 part Dye Reagent Concentrate with parts DDI water. This was then filtered through a Whatman #1 filter (or equivalent).

b. A Calibration curve was prepared by making a serial dilution of a 1mg/ml solution of BSA (10µg-1.25µg)

c. 10µl of test sample and standard was added to the plate.

d. 200 µl of diluted dye reagent was added to each well. The sample and reagent were thoroughly mixed using a multi-channel pipette.

e. The plate was incubated at room temperature for at least 5 minutes. Absorbance increases over time; samples should incubate at room temperature for no more than 1 hour.

f. The optical density of the samples was measured using a spectrophotometer at 590nm wavelength, using water as the blank.

g. Using the calibration curve, the protein concentration of the samples was determined using the computer software Grafit.

2.1.8.2 First Dimension separation by IEF

The first separation (first dimension) was carried out on a Multiphor II flat bed electrophoresis system (Pharmacia Biotech, Uppsala, Sweden). Ready made strips, Immobiline Drystrip gels (IPG) with a pH gradient 3-10 NL (non-linear) 13cm long (Amersham Pharmacia Biotech, Bucks, United Kingdom) were used. The strips were passively rehydrated overnight at room temperature with a rehydration buffer (8M Urea,

3% CHAPS, 0.5% IPG buffer 3-10, 10mM DTT, and a trace amount of bromophenol blue), which contained a volume equivalent of 75 μg of protein for preparative gels and 300 μg for analytical gels.. The IPG strips were over laid with 2 mls of IPG Cover Fluid (mineral oil) to minimise evaporation and urea crystallization. The rehydrated IPG strips were removed from the reswelling tray with a pair of forceps, rinsed with deionised water to remove excess rehydration solution (in order to prevent formation of urea crystals on the gel surface during IEF), and excess moisture drained off with filter paper. The IPG strips were then positioned in the DryStrip aligner with the gel side facing upwards in the grooves, with the acidic end near the anode. 11cm x 5mm electrode paper strips, moistened with distilled water, were then placed across the cathodic and anodic ends of the aligned IPG strips. The electrodes were then carefully inserted to contact the electrode strips, and about 50 ml of mineral oil poured onto the strips. IEF was carried out using multi-step conditions as follows: 150 V for 1 h, 300 V for 1 h, 1500 V for 1 h, and 3000V for a minimum of 18 h (Bio-Rad™ power pac 3000). The power supply was programmed in gradient mode with the current check option turned off. To establish cooling, the temperature on the Multitemp II Thermostatic Circulator was set at 20°C. The bromophenol blue tracking dye front migrates toward the anode as isoelectric focusing proceeds, but leaves the IPG strip before focusing is complete, therefore clearing of the dye doesn't imply that the sample is focused.

2.1.8.3 IPG Strip Equilibration

After the IEF run, the strips were equilibrated by incubating for 15 minutes at room temperature in 10 ml of equilibration buffer (50 mM Tris-HCl, pH 8.8, 6M urea, 30% w/v glycerol, 2% w/v SDS and a trace of Bromophenol Blue) with 1% w/v DTT. A second equilibrating step of 15 minutes was done in 10 ml equilibration buffer containing 4%w/v iodoacetamide (Gorg et al., 1998). Both solutions were freshly made prior to IPG strip equilibration. After equilibration, the IPG strips were aligned on filter paper moistened

with deionised water. Any excess equilibration solution was drained before the strips were applied to vertical SDS gels (Gorg et al., 2000).

2.1.8.4 Second Dimension SDS-PAGE

The second dimension separation was performed with home made 12.5%homogenous vertical SDS-polyacrylamide gel slabs (**Table 2.1**) (gel plate size 180 x 160mm, and gel thickness 1.5 mm) employing the tris-glycine system (Laemmli, 1970). The gel sandwich was prepared as per the Hoefer SE 600 series user manual. The gels were poured up to 0.5 cm from the top of the plates and overlaid with a thin layer (500 µL) of water-saturated with *n*-butanol, immediately after pouring to minimise gel exposure to oxygen and to create a flat gel surface. The focused and equilibrated IPG strips were then dipped in SDS electrophoresis buffer for lubrication and then transferred onto the upper edge of the SDS gel and sealed with agarose solution (Tris base 25 mM, glycine192 mM, SDS 0.1% w/v, agarose 0.5% w/v). Electrophoresis was performed at 20°C and a constant current of 25 mA per gel using SE 600 series vertical slab gel electrophoresis units. A constant current setting is traditionally used with a discontinuous buffer system so that the rate of electrophoresis migration remains unchanged throughout the run. Under these conditions, voltage increases as the run proceeds.

Table 2.1 Components of 12.5% SDS-PAGE gel used as second dimension

	2 gels	**4 gels**	**6 gels**	**8 gels**
Distilled Water	20 ml	40 ml	60 ml	80 ml
Acryl/Bis Soln	33 ml	66 ml	99 ml	132 ml
1.5M Tris (pH 8.8)	17ml	34 ml	51 ml	68 ml
TEMED	30 ul	60 ul	90 ul	120 ul
10% Ammonium persulphate in water (100mg/ml)	350 ul	700 ul	1203 ul	1400 ul

2.1.8.5 Detection techniques

Most detection methods used for SDS gels can be applied to second-dimension gels. Using staining procedures, proteins can be visualized as spots with varying properties or features, such as size, darkness/brightness and location on the gel. Silver staining and Coomassie staining were used in these investigations and the protocols used are described below.

Silver Staining

Analytical gels were silver stained using a protocol described by Blum (Blum et al., 1984) with modifications described by Rabilloud (Rabilloud et al., 1994b, Rabilloud, 1999). To fix the gels and remove any interfering compounds in the gel (glycine, tris and amopholytes) gels were placed in fixing solution (40% ethanol, 5% Acetic Acid, 55% distilled water) and left overnight. Gels were then washed in distilled water for 30 minutes. To increase the subsequent image formation, gels were sensitized with sodium hydrosulphite 0.3g/L of distilled water for 2 minutes. The gels were then washed 2x 1 min with distilled water. Silver impregnation was performed by soaking the gel in silver nitrate

2g/L of distilled water + 250 µl formaldehyde (added just before use) and stained for 40 min. Gels were rapidly washed in water for 20 seconds. The images were developed in 30 g Sodium Carbonate +250ul formaldehyde + 10mg of Sodium thiosulphate in 1L of water for 6-10 min depending on the desired intensity and before the background became too dark. The gels were then transferred to the stop solution (50g/L Tris base+ acetic acid 20ml/L) for a minimum of 2hr and then rinsed and left in deionised water.

Coomassie Staining

Preparative gels were stained using colloidal Coomasie Brilliant Blue G-250 as previously described (Neuhoff et al., 1988). Previously prepared Coomassie stock (**Table 2.2**) was used to make the stain by adding 1 part methanol to 4 parts of stock. The stain was left overnight or for a minimum of 2 hrs. The gels were then destained with 10% acetic acid in 25% methanol for 60 seconds. Using 25% methanol, gels were repeatedly rinsed until desired contrast was achieved. The gels were then left overnight in the 25% methanol. The methanol was changed for distilled water and stored at room temperature for scanning and spot analysis.

Table 2.2 Recipe for Colloidal Coomassie Stock

	2 Gels	4 Gels	6 Gels	8 Gels
10% w/v Ammonium Sulphate	40 g	80 g	120 g	160 g
2% v/v Orthophosphoric Acid	8 ml	16 ml	24 ml	32 ml
0.125% w/v CBB-G250	0.5 g.	1 g	1.5 g	2 g
Water	400 ml	800 ml	1200 ml	1600 ml

Scanning of Gel Image

The stained gels were scanned using GS-710 Imaging Densitometer (Bio-Rad). The gels were analysed using PDQuest® software version 6.2.1 (Bio-Rad) and Progenesis PG 220® V2006 software (Nonlinear Dynamics, Newcastle upon Tyne, UK), which enabled us to match identical spots in serial gels and normalise the gels to compensate for non-expression related variations in protein spot intensity.

Image analysis using PDQuest®

The work flow of PDQuest® comprises of the following steps following image acquisition by scanning:

(a) 2-DE gel image optimisation and editing which involves smoothing, contrast enhancement, edge detection and background subtraction. Background subtraction helps to eliminate meaningless changes of the background.

(b) *Protein spot detection:* is a multiple step procedure in which the gel image is processed and spot centres are marked according parameters one selects. The spots are then fitted to a Gaussian model. Gaussian fitted spots are more rounded and well defined compared to the original gel spots. Having a Gaussian model minimises having areas of overlapping and streaking allowing for more accurate quantitation, and Gaussian fitted spots are the ones used for higher level analysis and data in PDQuest®.

(c) *Gel matching* (comparison of identical spots in serial gels enabling comparative analysis of alterations in protein spot expression, under various experimental conditions): this is done by creating PDQuest® "matchsets". Matchsets are groups of gels with spots to be compared and they comprise of the following elements:

(i) *Matchset members*: which are gel spots images containing Gaussian images of gels created from the original scans. These Gaussian spots are used for matching and quantitation.

(ii) Member images: these are copies of the original gel scans and they contain the original spot data as well as the processing associated with spot data. One can toggle between the member with Gaussian spots and the member images of the original spots. The original scan files however, remain unchanged and are not included in the matchset.

(iii) Matchset standard, this is the synthetic gel image from the spots of the various members of a matchset generated by PDQuest®, that represents all spot data in the matchset. It is based on a template chosen from one of the match set members. The criteria for selecting the ideal template gel as the standard include: selecting a gel with a large

number of spots, one that is representative of the experiment as a whole, and in which the spots are well resolved and are of good quality. All matching and higher level calculations are performed using the standard.

(d) Data handling by generating replicate groups: Because of the complexity and inherent imprecision of the process of loading and running 2-DE gels, different gels loaded with the same samples may result in spots with differing amounts of proteins. To circumvent this hurdle, the same sample is run on multiple gels and spot data from these duplicate gels are combined for more accurate quantitation using PDQuest® replicate groups function. In this case the software requires that the duplicate groups are included in the same matchset in order to combine them in a replicate group. Using PDQuest® one is then able to show the average quantity of spots in each replicate group (replication group quantitation mode), or provide histograms showing the quantity the spot in each member (quantitation mode).

(e) *Normalisation of gels*: is a process of compensating for non expression related variation in protein spot intensity, important for accurate quantitation (PDQuest® uses house keeping proteins for this).

(f) *Gel data analysis using analysis sets*: involves grouping of proteins spots generated using PDQuest®, based on various statistical criteria (eg using student's t-test, Boolean analysis or higher level analysis based on two other analysis sets) or a manual selection of ones own design. Analysis sets are used to create and study categories of proteins of interest, through determination of qualitative and or quantitative changes in protein expression. Analysis sets are created and displayed in the match set standards.

(g) *Data presentation and interpretation*: involves assigning detected spots M_r and pI; comparison of spot quantitation and reflection of alteration trends of spots easily illustrated with graphs, scattergrams and bar charts.

(h) The analysis can be maintained in databases that link this information to the original gel images using PDQuest®, and 2-DE databases can be created.

Image analysis using Progenesis (PG 220)

The workflow of PG 220 comprises of the following steps following image acquisition by scanning:

a) Spot Detection: For Spot detection to be performed a reference gel of the experiment and average gels of the replicates are created.

   i) *Reference Gel:* Is an index of all the spots in an experiment and can be used in a single experiment or in several experiments. A reference gel is usually based on the slave (original) gel with the most spots.

   ii) *Averaged Gel:* Is an averaged representation of replicate gels of a sample type. One of the replicate gels is used as base for the average gel. The averaged gel takes into account replicate reproducibility and averaged gels spot data is represented as an average of matched spots. Each spot in an average is expressed as a mean with a standard deviation.

b) Background Subtraction: The software provides three automatic methods of background subtraction (mode of non spot, average on boundary, and lowest on boundary); in addition, a manual option is available.

   i) *Mode of non-spot*: This describes the modal average pixel value around the outside of a spot that does not belong to any neighbouring spot.

   ii) *Average on boundary*: uses the mean of all the pixels one pixel beyond the boundary of each spot

iii) *Lowest on boundary:* uses the lowest pixel value one pixel beyond the boundary of each spot

iv) *Multiple area selections (Manual)*: interpolates the background using the mean pixel value from a range of user-defined rectangles and the distance of a spot from each of these areas

c) Matching: A single reference gel may be matched to an unlimited number of gels in an experiment and between experiments (shared reference gels). Reference gels contain all the spots of interest through a variety of fully controllable automatic and semi-automatic methods. One can automatically match with "user defined seeds" to improve matching between very different or poorly run gels.

d) Warping: The image overlay warping function is used to align the reference gel spots to the spots of the current gel. Each additional user defined match (or seed) causes further localised warping. Real differences between gels can be instantly identified. After matching, the user can see the distortions in the gel using match vector lines, again a very useful visual aid.

e) Normalisation: This feature allows the user to normalise all gels in a series to known values in a gel, or to a value from which to draw comparisons between other proteins. This is vital when determining changes in expression across gels.

f) Protein Expression Queries: Once matching is complete a number of queries can be made as described below.

i) *Similarity Queries*: Find all the proteins that change in expression across a series of gels (or averaged gels) in the same way as any selected protein by simply clicking on a spot in the image.

ii) *Ratio Queries:* Define an expression ratio pattern across a series of gels (or averaged gels) and find all those proteins that fall within the limits of the defined ratio pattern.

iii) *Expression Query Parameters:* Highly flexible and fully visual parameter settings for broadening or limiting ratio definitions on a per gel basis.

i) Statistical Analysis: This tool consist of

i) *Differential analysis* - allows the user to identify all the series of matched spots that fit within definable ratio boundaries. This is ideal for identifying which proteins are within a set variation range across gels.

ii) *Student's t-test* - analyses sets of spots in averaged gel pairs in order to identify whether differences in measurements between matched spots are significant.

iii) *Correspondence analysis* - can aid the identification of spots that are characteristic of groups of gels. Using a scatter graph, gels that are close to each other on the graph are similar to one another, and in turn spots that are close to a group of gels are characteristic of those gels

2.2 One and Two Dimensional Liquid Chromatography

2.2.1 Back ground

Liquid chromatography (LC) is recognized as an indispensable tool in proteomics research since it provides high-speed, high-resolution and high-sensitivity separation of macromolecules. Chromatography entails the separation of a mixture of molecules dissolved in mobile phase. As the mobile phase travels over a stationary matrix, the components of the mixture can interact with the molecules of both the mobile phase and

the stationary matrix. An attractive feature of LC is the broad selection of stationary and mobile phases enabling a better separation of highly acidic and basic proteins as well as hydrophobic proteins which are hard to detect on 2-DE (Gygi and Aebersold, 2000). In addition, the unique features of chromatography enable the detection of low-abundance species making it an attractive alternative to 2-DE. Components such as phosphorylated proteins are often present in complex mixtures at extremely small concentrations.

A major difference between traditional high pressure liquid chromatography (HPLC) and the chromatography used in LC coupled to mass spectrometry (MS) is that in the latter the scale is usually much smaller, both regarding internal diameter of the column and even more so with respect to flow rate since it scales as the square of the diameter. For a long time 1 mm columns were standard for LC-MS (as opposed to 4.6 mm for HPLC) but more recently 300μm and even "capillary" 75μm columns have become more prevalent. At the low end of these column diameters the flow rates approach 100nL/min and are generally used with nanospray sources for mass spectrometry

Protein mixture analysis can be accomplished from one of two directions: top-down or bottom-up (Kettman et al., 2001). Top-down methods focus on the intact protein whereas bottom-up methods concentrate on the peptides that are usually produced from enzymatic digestion. Top-down methods, which include gel electrophoresis and liquid-based protein chromatography, separate proteins according to their molecular weight, isoelectric point, hydrophobicity, and/or binding affinity. These methods can support the resolution of protein states, such as the determination of phosphorylated and unphosphorylated forms of a protein.

In bottom-up methods (Wolters et al., 2001), the entire batch of proteins is digested to produce a very complex mixture of peptides. This is analogous to jigsaw puzzles, with a given puzzle representing one protein and its pieces representing the peptide components.

In bottom-up methods, all the pieces of all the puzzles are mixed together. The protein is identified by selecting and sequencing the individual peptides. When a high number of peptides are sequenced there is greater confidence that the correct protein is identified. Identification is the strength of bottom-up methods, but the weakness is full protein characterisation. Two similar proteins (i.e., phosphorylated and unphosphorylated) would have many of the same peptides. In bottom-up methods, high sequence coverage is rare, and characterising individual peptide modifications would be hit-and-miss depending on its selection from the massive pool.

Top-down analysis is capable of differentiating between two similar proteins as they are first separated and then analyzed. Unfortunately, the proteins cannot be identified without fragmentation either in solution (enzymatic digestion) or the gas phase. Newer techniques that allow gas-phase protein dissociation in the mass spectrometer support sequencing, but full sequence coverage is not yet possible for most proteins. The best route to fully characterise a protein involves protein separation and a multi-pronged digestion/sequencing scheme - a blend of top-down and bottom-up.

A variety of separation modes have been used both alone and in combination (Wang et al., 2003, Leinweber and Tallarek, 2003, Le Bihan et al., 2003) to achieve increased resolution of protein and peptides. The increased resolution reduces ion suppression effects in MS which are caused by overlapping signals from high and low abundance ions. Details of the application of LC in proteomic analysis are given below

2.2.2 S eparation Techniques in Proteomics

2.2.2.1 One dimension separation

Currently, most one dimension liquid chromatography (1D-LC) separations are done in reversed phase LC (RP-LC) mode because of its compatibility with MS. RP-LC consists of a non-polar stationary phase and a moderately polar mobile phase and involves the

reversible adsorption of proteins and peptides to the stationary phase matrix which are then eluted using solvent gradients in the mobile phase. RP-LC operates on the principle of hydrophobic interactions which result from repulsive forces between a relatively polar solvent, the relatively non-polar analyte, and the non-polar stationary phase. The driving force in the binding of the analyte to the stationary phase is the decrease in the area of the non-polar segment of the analyte molecule exposed to the solvent. This hydrophobic effect is dominated by the decrease in free energy from entropy associated with the minimisation of the ordered molecule-polar solvent interface. The hydrophobic effect is decreased by adding a more non-polar solvent into the mobile phase. This shifts the partition coefficient such that the analyte spends some portion of time moving down the column in the mobile phase, eventually eluting from the column.

The characteristics of the analyte molecule play an important role in its retention characteristics. In general, an analyte with a longer alkyl chain length results in a longer retention time because it increases the molecule's hydrophobicity. Very large molecules, however, can result in incomplete interaction between the large analyte surface and the alkyl chain. Retention time increases with hydrophobic surface area which is roughly inversely proportional to solute size. Branched chain compounds elute more rapidly than their corresponding isomers because the overall surface area is decreased.

Aside from mobile phase hydrophobicity, other mobile phase modifiers can affect analyte retention. For example, the addition of inorganic salts causes a linear increase in the surface tension of aqueous solutions, and because the entropy of the analyte-solvent interface is controlled by surface tension, the addition of salts tends to increase the retention time. Another important component is pH since this can change the hydrophobicity of the analyte. Organic acids such as formic acid (FA) or trifluoroacetic acid (TFA) are often added to the mobile phase. These serve multiple purposes: they control pH, neutralize the charge on any residual exposed silica on the stationary phase and

act as ion pairing agents to neutralize charge on the analyte. The effect varies depending on use but generally improve the chromatography. However, TFA concentrations are kept low if the column is coupled directly to the mass spectrometer as TFA has been known to cause ion suppression (Shi et al., 2004).

1D LC is economic but restricted to samples of minimal complexity and is often used for gel spot analysis. Analysis of complex protein mixtures using single dimension separation systems like the RPLC usually leads to poor chromatographic separation and coelution of peptides which leads to poor mass spectra. This compromises the analytical range of 1D-LC methods (Liu et al., 2002) and multi dimensional separations are often required.

2.2.2.2 Multiple dimensional Liquid Chromatography

Ion-exchange Chromatography-RPLC

In multiple dimensional liquid chromatography (2D-LC), RPLC is almost always preceded by ion-exchange chromatography (IEX) (Shi et al., 2004). In IEX chromatography, retention is based on the attraction between solute ions and charged sites bound to the stationary phase. Ions of the same charge are excluded. Some types of Ion Exchangers include: (1) Polystyrene resins- allows cross linkage which increases the stability of the chain. Higher cross linkage increases the equilibration time and ultimately improves selectivity. (2) Cellulose and dextran ion exchangers (gels) - these possess larger pore sizes and low charge densities making them suitable for protein separation. (3) And controlled-pore glass or porous silica. In general, ion exchangers favour the binding of ions of higher charge and smaller radius. An increase in counter ion (with respect to the functional groups in resins) concentration reduces the retention time. An increase in pH reduces the retention time in cation exchange while a decrease in pH reduces the retention time in anion exchange.

One use of IEX-RP is MudPIT (Multidimensional Protein Identification Technology) (Link et al., 1999, Wolters et al., 2001) which is a technique used for the separation and identification of complex protein and peptide mixtures. In this approach, the first dimension is normally a strong cation exchange (SCX) column, as these have high loading capacities. The second dimension is RP chromatography, which complements the SCX as it is efficient at removing salts and has the added advantage of being compatible with electrospray mass spectrometry.

Sample preparation is relatively straightforward, the samples are denatured, the cysteines reduced and alkylated, and the proteins digested with a protease such as trypsin. The samples are then acidified and loaded onto the SCX column. Charged peptides bind to the SCX column, whereas any uncharged peptides pass through and bind to a reverse phase trap column. The peptides are then eluted from the trap column onto an analytical RP column, using a reverse phase gradient, separated and eluted into a tandem mass spectrometer. Peptide fragmentation data is then obtained to identify the peptides and hence the proteins from which they are derived. In the next step, salt at a particular concentration is injected onto the SCX column, displacing further peptides from it onto the RP trap column. Salt is removed by washing and again an analytical RP separation is performed and the eluting peptides analysed by mass spectrometry. Incremental increases of salt are used (salt step gradient from around 0 - 200 mM). The end result is multiple protein identifications from each salt step.

Affinity Chromatography- RPLC

Modified proteins are sometimes present at very low concentrations making them difficult to detect and characterise in complex mixtures. Affinity chromatography is an effective and reproducible method to specifically isolate proteins and peptides. Immobilised metal affinity chromatography (IMAC) (Nuwaysir and Stults, 1993, Watts et al., 1994, Gaberc-

Porekar and Menart, 2001) utilises immobilised Fe (III) and Ga (III) ions to selectively retain phosphorylated proteins and peptides (Chaga, 2001). The method relies on high affinity interaction between transition metal ions and phosphorylated side chains of serine, threonine and tyrosine. Potential improvement of IMAC includes prevention of non-phosphorylated peptides to the IMAC column through carboxyl groups by converting peptides to methyl esters prior to analysis (Ficarro et al., 2002). Other improvements can be made by using strong anion exchange chromatography prior to IMAC to reduce complexity of IMAC purified phosphopeptides (Nuhse et al., 2003).

Immunoaffinity chromatography (IAC) can also be used to isolate and purify phosphorylated proteins prior to RPLC. In this technique the antibody is immobilized on the column resin. Peptide digests of phosphoproteins are loaded onto the column followed by washing and selective elution to antigenic peptides (Shi et al., 2004). IAC is highly selective due to high affinity of antigen-antibody interaction (Burgess and Thompson, 2002).

Size Exclusion Chromatography-RPLC

In size exclusion chromatography, the column is packed with inert beads made of a porous compound such as agarose. Large proteins find a quicker path through the column where small proteins take longer. There is no chemical interaction between the solute and the stationary phase. This method can be utilised for protein separation prior to tryptic digestion. The resulting peptide mixture is then separated by RPLC prior to mass spectrometry.

2.2.3 E xperimental Protocols

2.2.3.1 Sample Preparation

Tryptic in-gel digestion

Eppendorf tubes and all utensils were cleaned with 50% v/v acetonitrile/0.1% v/v TFA solution and dried. The protein spots of interest identified on preparative gels were excised and transferred to the eppendorf tubes. The chopped gel pieces were washed in 50% acetonitrile/ 25mM ammonium bicarbonate, pH 7.8, and dried in a vacuum concentrator. 4-10 µl digestion buffer (10 µg/ml modified sequencing grade trypsin (Sigma-Aldrich) in 25 mM ammonium bicarbonate) was added to the dried gel pieces and incubated overnight at 37°C. Resulting peptides were extracted by addition of 4µl water followed by 7µl of 30% acetonitirile/0.1% TFA followed by vortexing and brief centrifugation. The supernatant was transferred to a clean tube and vacuum concentrated to approximately 5µl.

In Solution Tryptic Digests

Samples were reduced in 3µl of 0.01M DTT in water for every 10µg of protein. This was done for 30 minutes at 30°C. To prevent the disulphide bonds from reforming, samples were then alkylated using 3µl of 0.01M IAA in 50mM NH_4HCO_3 for every 10µg of protein and incubated for another 30 minutes at room temperature (~20°C-22°C). To quench any excess IAA an excess of the 0.01M DTT was added to the solution. 1µg (20µg/ml) of trypsin was added for every 100µg of protein to the sample and incubated overnight at 37°C.

2.2.3.2 Offline HPLC Fraction Collection

The method described here was used to reduce the complexity of the samples prior to proteolytic enzyme digestion. The HPLC was performed on an UltiMate™ 3000 LC system (Dionex UK). Samples were separated on a ProSwift RP-1S™monolith column (4.6 × 50 mm Dionex UK). Solvent A was 2.5% Acetonitrile (ACN) in water with 0.1% TFA and solvent B was 90% ACN in water with 0.1% TFA. The flow rate was maintained at 200µl/minute and the proteins eluted into a 396 well plate using an ACN gradient.

Fractions were collected every half a minute from one minute to 25 minutes resulting in 48 fractions collected per sample.

2.2.3.3 Online Second Dimension

The method described here was used to separate tryptic peptides from 2DE gel spots and digested LC fractions. Samples were injected onto a Monolithic Capillary Column, 200 μm i.d. x 5 cm (Dionex UK) or Thermo Hypersil-Keystone biobasic C18 column (0.18 x 100mm) and peptides were eluted to the mass spectrometer using various isocratic gradients depending on the column used and the type of sample. The gradients will be described in the respective chapters.

2.3 Mass Spectrometry for Proteomics

2.3.1 Background

Mass spectrometry is an analytical technique used to measure the mass-to-charge ratio (*m/z*) of ions. It is most generally used to find the composition of a physical sample by generating a mass spectrum representing the masses of sample components. A typical mass spectrometer comprises three parts: 1) an ion source-in which ions are produced from the sample, 2) a mass analyzer-which separates the ions according to their *m/z*, and 3) a detector system- which is usually an electron multiplier that produces a signal proportional to the relative abundance of each ion resulting in the number of ions produced by each mass. The three parts are not complete without a data system which controls the operation of the instrument as well as process and display data in a mass spectrum. The technique has several applications, including: identifying unknown compounds by the mass of the compound molecules or their fragments; determining the isotopic composition of elements in a compound; determining the structure of a compound by observing its fragmentation; quantifying the amount of a compound in a sample using carefully designed methods (mass spectrometry is not inherently

quantitative); studying the fundamentals of gas phase ion chemistry (the chemistry of ions and neutrals in vacuum); determining other physical, chemical, or even biological properties of compounds with a variety of other approaches. Details of application of this technique in proteomics are given below.

2.3.2 Ionisation Techniques

The ion source is the part of the mass spectrometer that ionizes the material under analysis (the analyte). The ions are then transported by magnetic or electric fields to the mass analyzer. Techniques for ionisation have been key to determining what types of samples can be analysed by mass spectrometry. Electron ionisation and chemical ionisation are used for gases and vapours. In chemical ionisation sources, the analyte is ionised by chemical ion-molecule reactions during collisions in the source. Two techniques often used with liquid and solid biological samples include electrospray ionisation (ESI) (Fenn et al., 1989) and matrix-assisted laser desorption/ionization (MALDI) (Hillenkamp et al., 1991, Koichi Tanaka, 1988, Karas et al., 1987). Inductively coupled plasma sources are used primarily for metal analysis on a wide array of samples types. Other techniques include glow discharge, fast atom bombardment (FAB), thermospray, desorption/ionisation on silicon (DIOS), Direct Analysis in Real Time (DART), atmospheric pressure chemical ionization (APCI), secondary ion mass spectrometry (SIMS) and thermal ionisation. Studies described in this thesis have used MALDI and ESI.

2.3.3 *Matrix-Assisted Laser Depsorption and ionisation (MALDI)*

In MALDI, the sample to be analysed is dissolved in a matrix, usually an aromatic acid such as α-cyano-4-hydroxy-cinnamic acid or sinapinic acid. A sample-matrix mixture is then spotted onto a target plate and allowed to dry. The matrix incorporates the biopolymers into the crystals formed as it dries. The plate is then placed into the machine, and the sample is ionised and vaporised with a pulsed nitrogen laser at 337nm. This leads

to the sputtering of pulses of MH^+ ions from each compound in the sample and these ions are then introduced to the MS. MALDI has a useful mass range of over 250,000 Daltons (Da) and is very tolerant of salts, buffers and detergents.

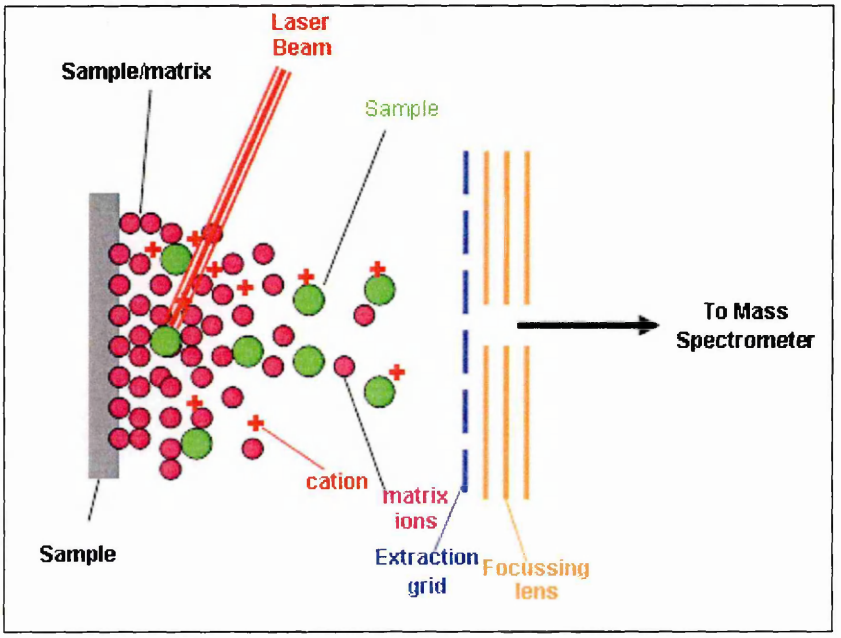

Figure 2.1 A schematic diagram of the mechanism of MALDI

2.3.4 Electrospray ionisation (ESI)

Electrospray is a method where a liquid is dispersed into small charged droplets by applying a high electric potential between a liquid in a thin capillary and a counter electrode. The charging of the analytes normally takes place in the liquid phase, whereupon the ions are transferred to the gas phase. The charging of the analytes can occur in different ways: the analytes may already be charged in solution, by adduct formation, gas-phase ionisation or electrochemical ionisation. Peptides may attach to one or more protons so that they give rise to ions such as $(M+H)^+$, $(M+2H)^{2+}$, $(M+3H)^{3+}$. Proteins may attach many more protons and give rise to a large number of ions of type $(M+nH)^{n+}$. The upper mass range for ESI is uncertain but ions in excess of 2,000,000 have been reported

(Siuzdak, 1996). It is less tolerant of salt, buffers and detergents than MALDI and it is essential to remove these for good results.

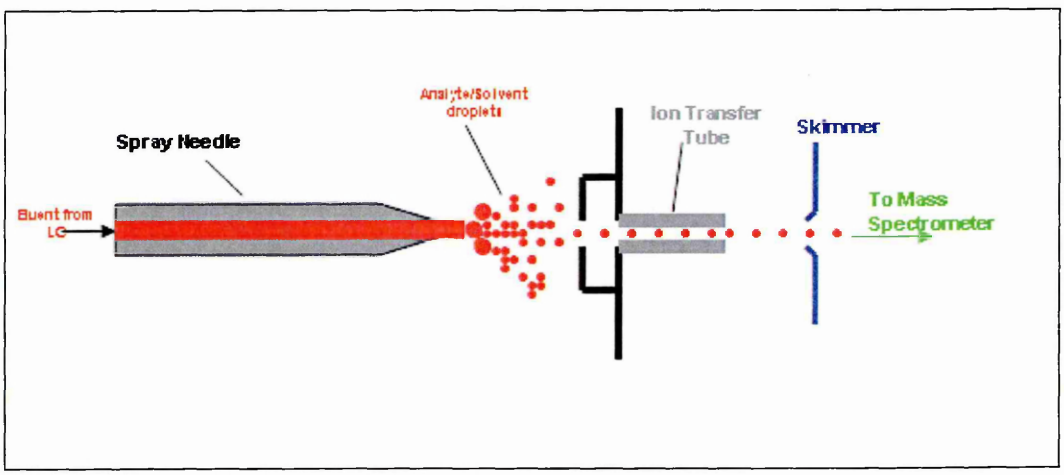

Figure 2.2 A schematic of an ESI source.

2.3.5 Mass Analysers

All commonly used mass analyzers use electric and magnetic fields to apply a force on charged particles (ions). The main types of analysers currently in use are magnetic sector, time-of-flight (ToF), quadrupole, ion trap and fourier transform ion cyclotron resonance (FTICR).

2.3.5.1 Magnetic Sector Mass Analyser

In a magnetic deflection mass spectrometer, ions leaving the ion source are accelerated to a high velocity. The ions then pass through a magnetic sector in which the magnetic field is applied in a direction perpendicular to the direction of ion motion. A magnetic sector alone will separate ions according to their mass-to-charge ratio. However, the resolution will be limited by the fact that ions leaving the ion source do not all have exactly the same energy and therefore do not have exactly the same velocity. These analysers are used less frequently than ToF, quadruple and ion trap as they are less sensitive.

2.3.5.2 Time-of-flight (ToF) Mass Analyser

A time of flight mass spectrometer measures the mass-dependent time it takes ions of different masses to move from the ion source to the detector. This requires that the starting time (the time at which the ions leave the ion source) is well-defined. Therefore, ions are either formed by a pulsed ionization method (usually MALDI), or various kinds of rapid electric field switching are used as a 'gate' to release the ions from the ion source in a very short time. Ions in the source are subjected to a pulse of about 25-30kV so that in principle they all have the same transitional energy ($1/2mv^2$, where m is the mass and v is the velocity) hence the lightest ions will travel fastest and reach the detector at the end of the tube in the shortest time. The recorded time of flight (ToF) measurements are used to calculate the masses of ions in m/z. Separation of ions is best at low m/z and the resolving power falls as m/z rises.

In theory, all ions experience the same potential difference during acceleration and thus have the same kinetic energy at the start of the flight tube, and thus different velocities depending upon their mass. Therefore, their arrival time at the detector is proportional to their mass and they reach the detector in order of increasing mass. The mass-to-charge ratios (m/z) of the ions relate to the flight times t by the following equations:

$mv^2/2=zV;\ v=\sqrt{2zVm}$ (Equation 1)

$t=l/v=l\sqrt{m/2zV}$ (Equation 2)

where l is the length of the flight tube, v is the velocity of the ion of mass m and charge z, and V is the acceleration potential. The differences of flight times are the basis for resolving ions of different m/z in the ToF analyzer and the mass resolution depends on the flight time differences, which are proportional to $l(\sqrt{m}_1 - \sqrt{m}_2)$.

The major deficiency of a simple linear ToF instrument is its insufficient mass resolution resulting from flight time variations of ions of the same *m/z*. First, the laser pulses used for ion generation in MALDI-ToF instruments are fairly long, resulting in a wide distribution of starting times of the same ions. Further, the ionization process adds a certain amount of initial kinetic energy to the molecules before acceleration. Finally, the different spatial positions in the source from where the ions are formed lead to varying values of *l* (Equation 2) and thus leading to flight time variations. As a result, modern ToF instruments commonly employ two techniques to enhance resolution-reflectron and delayed extraction.

A reflectron ToF is used to focus ions of the same *m/z* but of different kinetic energy. There are different designs of reflectrons but often they consist of a series of rings or grids. The reflectron is located after the drift tube, creating a retarding field that the ions penetrate. Depending upon their kinetic energy, they enter this field at different depths and then are reflected back into the flight tube, where they drift to the detector, which is placed close to the ion source. Consider the case of ions of the same *m/z* but significant energy spread, which would be detected as broad peaks in the linear ToF: the ion with a higher kinetic energy penetrates deeper and spends more time in the reflectron. The slow ion, on the other hand, returns to the flight tube faster, but with the same lower kinetic energy. The faster ion will catch up with the slow one at the detector, thereby improving mass resolution greatly. However, this increased resolution comes at the expense of sensitivity and has a limited mass range (Siuzdak, 1996).

2.3.5.3 Quadrupole Mass Analyser

The quadrupole is the most common analyzer found in analytical laboratories today. It is not only the standard analyzer for GC–MS (which will not be discussed here) but also is part of various designs of LC–MS instruments. Quadrupoles are not used only as mass

analyzers; they also are implemented frequently as ion transfer optics, collision cells, and linear ion traps (*vide infra*). This mass analyzer works on the basis of electric fields generated between a set of four axial rods through which ions pass on their way to the detector. The voltages applied to these rods consist of direct current (DC) and radio frequency (RF) components, which together create a quadrupolar field, allowing only a certain *m/z* range to pass through for a given combination of potentials; ions outside that *m/z* region hit the rods and are discharged. The quadrupole therefore is referred to also as a *mass filter*, into which ions are accelerated from the ion source by a small potential. Positive ions will be attracted to negative rods, and vice versa. The RF component, however, causes the electric field to alternate at a frequency in the megahertz range. The applied potential to the rods is of the form:

$\Phi_0 = U + V\cos \omega t$

where U is the DC voltage and $V\cos \omega t$ is the RF potential of frequency $\omega/2\pi$. Two RF waveforms are applied to two pairs of opposing parallel rods, with a 180° phase shift. As a result, the trajectory of an ion fluctuates constantly as it travels between the rods until its detection.

2.3.5.4 Ion Trap

The ion-trap mass spectrometer may be looked upon as a three-dimensional quadrupole mass analyzer that uses three electrodes to trap ions in a small volume. The mass analyzer consists of a ring electrode separating two hemispherical electrodes. A mass spectrum is obtained by changing the electrode voltages to eject the ions from the trap. The advantages of the ion-trap mass spectrometer include compact size, and the ability to trap and accumulate ions to increase the signal-to-noise ratio of a measurement.

Ions are formed within the ion trap or injected into an ion trap from an external source. The ions are dynamically trapped by the applied RF potentials (a common trap design also makes use of a "bath gas" to help contain the ions in the trap). The trapped ions can be manipulated by RF events analogous to the events in FTICR to perform ion ejection, ion excitation, and mass-selective ejection. This provides MS/MS and MS/MS/MS experiments analogous to those performed in FTICR.

Space-charge effects (ion-ion repulsion) severely limit the inherent dynamic range of the ion trap. This is usually handled by auto-ranging where a pre-scan is performed to determine the ion current, and then the ionising electron current is adjusted to reduce the number of ions formed to within the working range. This can be done wherever the ion formation event can be manipulated to control the number of ions formed (such as in electron ionization). In normal operation, a high scan rate is used to improve sensitivity but this comes at a cost of reduced resolving power. A typical commercial ion trap has a mass range of up to 2,000 m/z at normal resolving power but this can be extended to 6,000 m/z at reduced resolving power.

2.3.5.5 Fourier Transform-ion cyclotron Resonance (FTICR)

This mass analyzer determines the mass-to-charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field. The ions are trapped in a Penning trap (a magnetic field with electric trapping plates) where they are excited to a larger cyclotron radius by an oscillating electric field perpendicular to the magnetic field. The excitation also results in the ions moving in phase (in a packet). The signal is detected as an image current on a pair of plates which the packet of ions passes close to as they cyclotron. The resulting signal is called a free induction decay (FID), transient or interferogram that consists of a superposition of sine waves. The useful signal is extracted from this data by performing a Fourier transform to give a mass spectrum.

FTICR is a very high resolution technique in that masses can be determined with very high accuracy. Many applications of FTICR-MS use this mass accuracy to help determine the composition of molecules based on accurate mass. This is possible due to the mass defect of the elements. Another place that FTICR-MS is useful is in dealing with complex mixtures since the resolution (narrow peak width) allows the signals of two ions of similar m/z to be detected as distinct ions. This high resolution is also useful in studying large macromolecules such as proteins with multiple charges which can be produced by electrospray ionization. These large molecules contain a distribution of isotopes that produce a series of isotopic peaks that are close to each other on the m/z axis, due to the multiple charges, making the high resolving power of the FTICR extremely useful in this instance.

FTICR-MS differs significantly from other mass spectrometry techniques in that the ions are not detected by hitting a detector such as an electron multiplier but only by passing near detection plates. Additionally the masses are not resolved in space or time as with other techniques but only in frequency. Thus, the different ions are not detected in different places as with sector instruments or at different times as with time-of-flight instruments but all ions are detected simultaneously over some given period of time.

2.3.6 Peptide Mass Fingerprinting (PMF)

PMF is an analytical technique for protein identification that was developed by John Yates and colleagues (Link et al., 1999). In short, the unknown protein of interest is cleaved into peptides by a protease such as trypsin. The collection of peptides resulting from this cleavage, comprise a unique identifier of the unknown protein. The absolute masses of the (still unknown) peptides are accurately measured with a mass spectrometer such as MALDI-ToF or ESI-ToF. These masses are then compared to the genome. Computer programs translate the known genome of the organism into proteins, then theoretically cut the proteins into peptides with the same protease (for example trypsin), and calculate the

absolute masses of the peptides from each protein. Then they compare the masses of the peptides of the unknown protein to the theoretical peptide masses of each protein encoded in the genome. The results are statistically analysed to find the best match. Algorithms allowing database searching on the basis of peptide mass data were developed simultaneously in the early 1990s and have been incorporated in several software packages including Bioworks ™ (Thermo Electron), which uses the SEQUEST algorithm developed in Yates' laboratory (Link et al., 1999) and MACOT (Perkins et al., 1999) ™ (http://www.matrixscience.com)

The great advantage of PMF is that only the masses of the peptides have to be known (so *de novo* sequencing is not necessary). A disadvantage is that the protein sequence has to be present in the database of interest. Additionally most PMF algorithms assume the peptides come from a single protein. The presence of a mixture can significantly complicate the analysis and potentially compromise the results.

2.3.7 Fragment Ion Analysis Using Tandem Mass Spectrometry

Tandem mass spectrometry involves multiple steps of mass selection or analysis, usually separated by some form of fragmentation. A tandem mass spectrometer is capable of multiple rounds of mass spectrometry. For example, one mass analyzer can isolate one peptide from many entering a mass spectrometer. A second mass analyzer then stabilizes the peptide ions while they collide with a gas, causing them to fragment by CID. A third mass analyzer then catalogues the fragments produced from the peptides. Tandem MS can also be done in a single mass analyzer over time as in a quadrupole ion trap. There are various methods for fragmenting molecules for tandem MS, including collision-induced dissociation (CID), electron capture dissociation (ECD), electron transfer dissociation (ETD), infrared multiphoton dissociation (IRMPD) and blackbody infrared radiative dissociation (BIRD). An important application using tandem mass spectrometry is in protein identification.

In our studies we have used the ion trap to carry out tandem mass spectrometric experiments as follows. All ions, regardless of mass, are allowed to enter the ion trap and retained. A full scan (MS) generates a spectrum displaying the ion count of each mass initially entering the trap. Ions in the trap are exposed to altering voltages and a window of ions isolated, thus retaining only ions of interest (parent ions) which have a precursor mass. These ions are then subjected to CID and fragment/daughter ions are generated. Peptides subjected to CID fragment predominantly along their amide backbone. Fragments containing N-teminus are denoted *a, b* and *c* using the nomenclature described by Biemann (Biemann, 1988). C-terminus ions are denoted as *x, y* and *z*.

The most informative fragments are those in which breakage has occurred along the polypeptide backbone because this represent strings of contiguous and intact amino acids (Twyman, 2004). If the charge remains on the N-terminus it is known as a b-series ion and if is on the C-terminus it is known as a y-series ion.

2.3.8 Bioinformatics for Database Searching and Protein Cataloguing

Bioinformatics is a rapidly growing field which is inextricably linked with protein and peptide mass spectrometry. Bioinfromatic tools are normally used to scan databanks of sequence information to find potential matches to experimental data. Databases are collated, stored and managed by teams of experts such as the National Centre for Biotechnology (NCBI), the Sanger Centre and the European Bioinformatics Institute (EBI). Choice of database is dependent on a number of factors including the species one is working with, whether that species is well characterised and whether the proteins are known or unknown. Most database searching programs will allow users to adjust search parameters to suit the data being cross-correlated. Common search parameters include mass lists generated from the MS; choice of protease used so as to theoretically digest the database of choice; database species of choice; expected modifications such as phosphorylation (full list on **Table 2.3**);

Table 2.3 List of Common Post Translational Modifications

Sulphydryls	Disulphide bond	-2.0159	Oxidation	15.9994
(C)	Cysteinylation	119.1442	Glutathionylation	305.3117

Amines	Methylation	14.0269	Formylation	28.0104
(K/N-terminus)	Acetylation	42.0373	Lipoic acid	188.3147
	Farnesylation	204.3556	Myristoylation	210.3598
	Biotinylation	226.2994	Palmitoylation	238.4136
	Stearoylation	266.4674	Geranylgeranylation	272.4741

Acids & amides	Pyroglutamic acid (Q)	-17.0306	Deamidation (Q/N)	0.9847
(E/D/Q/N)	Carboxylation (E/D)	44.0098		

Hydroxyl groups	Phosphorylation	79.9799	Sulphation	80.0642
(S/T/Y)				

Carbohydrates	Pentoses	132.1161	Deoxyhexoses	146.143
(S/T/N)	Hexosamines	161.1577	Hexoses	162.1424
	N-acetylhexosamines	203.195	Sialic acid	291.2579

Different databases use different vocabularies to describe the function of proteins which has led to a need for an efficient and comprehensive tool for the annotation of functional genome information. To harvest the full power of protein identities, several functional cataloguing tools have been created to generate standardised functional vocabulary used. The objective of these tools is to associate gene products and functional data beyond the often inconsistent naming conventions used (Ruepp et al., 2004). An example of these tools include those used at the Protein Information Resource (PIR) (Barker et al., 2000), SWISS-PROT (Bairoch and Apweiler, 2000), and FunCat (Ruepp et al., 2004), to name a few. In this study we have used a tool developed by PIR which integrates the PIR super family (PIRSF) system, *a network classification system based on evolutionary relationship of full-length proteins*, to facilitate the propagation and standardisation of protein

annotation. Sequence analysis and protein classification based on full-length proteins, including the preservation of domain architecture, can lead to educated predictions for both generic biochemical and specific biological functions. The multiple levels of sequence diversity, from superfamilies to subfamilies, reflect different degrees of functional granularity and, thereby, allow more accurate propagation of annotation and the development of standard protein nomenclature and ontology. The PIRSF database is integrated with other family, function, and structural classification schemes, and is accessible at http://pir.georgetown.edu/pirsf/ for report retrieval and sequence classification.

2.3.8.1 Gene Ontologies (GO)

The PIR tool integrates gene ontology (GO) terms which were created as part of the Gene Ontology project formed as a collaborative effort to address the need for consistent descriptions of gene products in different databases. The project began as collaboration between three model organism databases, FlyBase (Drosophila), the Saccharomyces Genome Database (SGD) and the Mouse Genome Database (MGD), in 1998. Since then, the GO Consortium has grown to include many databases, including several of the world's major repositories for plant, animal and microbial genomes.

There are three separate aspects to the creation of GO: first, the development and maintenance of the ontologies themselves; second, the annotation of gene products, which entails making associations between the ontologies and the genes and gene products in the collaborating databases; and third, development of tools that facilitate the creation, maintenance and use of ontologies.

The use of GO terms by collaborating databases facilitates uniform queries across them. The controlled vocabularies are structured so that they can be queried at different levels:

for example, one can use GO to find all the gene products in the mouse genome that are involved in signal transduction, or one can examine all the receptor tyrosine kinases. This structure also allows annotators to assign properties to genes or gene products at different levels, depending on the depth of knowledge about that entity.

The GO project has developed three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. A gene product might be associated with or located in one or more cellular components; it is active in one or more biological processes, during which it performs one or more molecular functions. For example, the gene product cytochrome c can be described by the molecular function term oxidoreductase activity, the biological process terms oxidative phosphorylation and induction of cell death, and the cellular component terms mitochondrial matrix and mitochondrial inner membrane.

Cellular component

A cellular component may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer).

Biological process

A biological process is series of events accomplished by one or more ordered assemblies of molecular functions. Examples of broad biological process terms are cellular physiological process or signal transduction. Examples of more specific terms are pyrimidine metabolism or alpha-glucoside transport. It can be difficult to distinguish between a biological process and a molecular function, but the general rule is that a process must have more than one distinct steps.

A biological process is not equivalent to a pathway and GO does not try to represent the dynamics or dependencies that would be required to fully describe a pathway.

Molecular function

Molecular function describes activities, such as catalytic or binding activities, that occur at the molecular level. GO molecular function terms represent activities rather than the entities (molecules or complexes) that perform the actions, and do not specify where or when, or in what context, the action takes place. Molecular functions generally correspond to activities that can be performed by individual gene products, but some activities are performed by assembled complexes of gene products. Examples of broad functional terms are catalytic activity, transporter activity, or binding; examples of narrower functional terms are adenylate cyclase activity or Toll receptor binding. It is easy to confuse a gene product name with its molecular function, and for that reason many GO molecular functions are appended with the word "activity".

2.3.8.2 The MASCOT® Search Tool

The MASCOT® search engine is based on a modified MOWSE computer program (Pappin et al., 1993) using a probability based scoring system-the probability that the observed match between the experimental data and a protein sequence is random and is calculated for each protein sequence in the database. The proteins are classified with increasing probability of being a random match to the experimental data. The probability is based on the size of the database and the frequency of each peptide mass in each database. The search tool guidelines state that if the probability score is higher than the database probability *(p)* then the match is significant within 95% confidence limit. Results are displayed as a histogram showing the distribution of the scores and the 95% confidence limit.

2.3.8.3 The TurboSEQUEST® /SEQUEST® Search Tool

The TurboSEQUEST® algorithm in BioWorks™ v 3.1(Thermo Fisher) is based on the SEQUEST® algorithm (Eng et al., 1994, Yates et al., 1995a, Yates et al., 1995b). This algorithm enables one to cross-correlate un-interpreted MS/MS mass spectra of peptides with single-letter coded amino acid sequences from protein databases or with 3 letter coded nucleotide sequences from nucleotide databases. Both versions determine the amino acid sub-sequence and thus the protein(s) and organism(s) that correspond to the mass spectrum of the peptide being analyzed. Criteria for the database search are the molecular weight (MW) information of the peptide and the MS/MS data.

2.3.8.4 The NCBInr Database

This database contains the translated protein sequences from the entire collection of annotated DNA sequences kept at GenBank, and also protein sequences in the protein Data Bank (PDB), CDS translations, Protein Research Foundation (PRF), SWISS-PROT and Protein Information Resource (PIR) database, covering most of the publicly available data. The NCBI also provide BLAST (Basic Logical Alignment Search Tool), a program support for database sequence similarity searching

2.3.8.5 The Plasmodium Falciparum Database

The local *P. falciparum* database was constructed by downloading all available sequence information from the ftp site of the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium/falciparum/3D7/). This database includes all submitted sequences from GeneDB (Hertz-Fowler et al., 2004), which includes sequences from Ensembl, a joint project between The European Molecular Biology Laboratory (EMBL)- European Bioinformatics Institute (EBI), and the Wellcome Trust Sanger Institute (WTSI).

2.3.9 Experimental Protocols

2.3.9.1 MALDI-ToF Mass Spectrometry

MALDI peptide mass fingerprinting was carried out using a AXIMA CFR Plus, (Shimadzu Biotech-Kratos Analytical, Manchester UK) operating in positive ion reflectron mode at an accelerating voltage of 20kV. The spectra were externally calibrated using a peptide mixture (Sigma, St Louis, MO) with masses 757.39 (Bradykinin), 1046.54 (Angiotensin II) and 2465.19 (ACTH). Protein identification was performed using the MASCOT® search engine (Matrix Science, London, UK).

2.3.9.2 ESI-Ion Trap Tandem Mass Spectrometry

Tandem mass spectrometry was carried out on a LCQ Deca XP Plus ion trap mass spectrometer (ThermoFinnigan, USA) equipped with an electrospray source. The mass spectrometer was set at different settings depending on whether the source was in microspray or nanospray configuration and this will be fully described in the respective chapters. The LCQ Deca XP Plus was tuned using Angiotensin I (singly charges at m/z 1296.7, doubly charged at m/z 648.8 and triply charged at m/z 432.9) and calibrated according to the manufacturers' instructions. The spectra were evaluated using the TurboSEQUEST® algorithm in BioWorks ™v 3.1 software provided by ThermoFinnigan.

2.3.9.3 Protein identification Using MASCOT®

The workflow for MASCOT® comprises of these steps following peptide mass fingerprinting:

a. *Selecting a mass list*

A mass list is copied from the spectra generated by the MALDI-ToF and pasted in to the data query window in MASCOT®. The mass tolerance is set at ±1Da

b. *Choosing an enzyme*

Specify the reagent used for protein digestion. Table 2.4 gives a list of supported enzymes.

Table 2.4 A List of Enzymes supported by MASCOT®

Name	Cleave	Don't cleave	N or C term
Trypsin	KR	P	CTERM
Arg-C	R	P	CTERM
Asp-N	BD		NTERM
Asp-N_ambic	DE		NTERM
Chymotrypsin	FYWL	P	CTERM
CNBr	M		CTERM
Formic_acid	D		CTERM
Lys-C	K	P	CTERM
Lys-C/P	K		CTERM
PepsinA	FL		CTERM
Tryp-CNBr	KRM	P	CTERM
TrypChymo	FYWLKR	P	CTERM
Trypsin/P	KR		CTERM
V8-DE	BDEZ	P	CTERM
V8-E	EZ	P	CTERM
CNBr+Trypsin	M		CTERM
	KR	P	CTERM
None	see notes		
semiTrypsin	see notes		

NB: "None" means that MASCOT will search each protein sequence for every sub-sequence which meets the other search criteria. This may mean testing several orders of magnitude more peptides than if trypsin cleavage had been specified. If the experimental data are from peptides which do not originate from an enzyme digest, such as MHC peptides, then "None" is the correct choice. Otherwise, ignoring the enzyme specificity increases the number of random matches, which will reduce scoring discrimination. "None" is not an allowed choice for a Peptide Mass Fingerprint, where the specificity of an enzyme is essential. "semiTrypsin" means that MASCOT® will search for peptides that show tryptic specificity (KR not P) at one terminus, but where the other terminus may be a non-tryptic cleavage. This is a half-way house between choosing "Trypsin" and "None". It will only fail to find the rare peptides that are the product of double non-specific cleavage

c. *Assigning number of mass cleavages allowed*

Setting the number of allowed missed cleavage sites to zero simulates a limit digest. This would only be used if confident that a digest is perfect, with no partial fragments present. This will give maximum discrimination and the highest score.

Our digest mixtures usually include some partials, that is, peptides with missed cleavage sites, and setting 2 missed cleavage sites is normally used.

d. *Choosing a database*

The databases available on the MASCOT® server include EST (EST divisions of Genbank, currently EST_human, EST_mouse, EST_others), MSDB (a comprehensive, non-identical protein database) NCBInr (a Comprehensive, non-identical protein database) and SwissProt (a high quality, curated protein database. For a Peptide Mass Fingerprint, the EST databases are not available because the entries are just short stretches of sequence, not complete proteins.

e. *Selecting Taxonomy*

The Taxonomy parameter allows searches to be limited to entries from particular species or groups of species. This can speed up a search, and ensures that the hit list will only contain entries from the selected species. If the search data are marginal, and you are completely confident of the origin of the protein, this can help bring a weak match to the top of the list.

f. *Setting modifications*

Select any known or suspected modifications. MASCOT® supports two types of modification. Fixed modifications are applied universally, to every instance of the specified residue(s) or terminus. There is no computational overhead associated with a fixed modification, it is simply equivalent to using a different mass for the modified residue(s) or terminus. For example, selecting Carboxymethyl (C) means that all calculations will use 161 Da as the mass of cysteine.

Variable modifications are those which may or may not be present. MASCOT® tests all possible arrangements of variable modifications to find the best match. For example, if Oxidation (M) is selected, and a peptide contains 3 methionines, MASCOT will test for a

match with the experimental data for that peptide containing 0, 1, 2, or 3 oxidised methionine residues.

2.3.9.4 Searching Using TurboSEQUEST® /SEQUEST®

The workflow for SEQUEST/TurboSEQUEST comprises of the following steps after MS/MS mass spectra acquisition.

a. Generating data files

A data (DTA) file contains the calculated singly protonated peptide mass (MH+) and the peptide charge state as a pair of separated values. If a zoom scan is present the protonated peptide mass is calculated from the zoom scan. Subsequent lines in the DTA file contain space separated pairs of fragment ion m/z and intensity values which come from the MS/MS data of the peptide mass (MH+).

A separate DTA file is created for every MH+ peptide mass in a single LCMS experiment.

1. Theoretical Database digestion

A theoretical digestion of the database using an appropriate enzyme is performed. The generated theoretical peptides are then mass sorted and for each peptide, a theoretical list of *b* and *y* ions is predicted to generate a series of 'virtual spectra.'

2. Searching the DTAs

Each DTA file created from the LCMS experiment is searched against the theoretical 'digested' database. All peptide candidates from the database with the same MW as our DTA ± a specified mass tolerance is now compared to the "theoretic MS^2 spectrum" of each of the database peptide candidates selected by MW. Each peptide candidate is given a primary match (Sp). The scoring is based on the number of *b* and *y* ions in the theoretical MS^2 spectrum of the peptide candidate which match the *b* and *y* from the DTA. Larger peptides and "noisy" spectra tend to have bigger Sp values as ion intensities are not taken into account. Each candidate is then given a relative

primary score (RSp) from the best match (250-500 peptide candidates are normally considered). The XCorr value is the cross-correlation value from the search. The #1 hit will always have the highest value of Xcorr, as Xcorr is used to produce the final ranking of the candidate peptides in the search. XCorr values above 2.0 are usually indicative of a good correlation. However, as with Sp, XCorr values are usually higher for well-matched, large peptides, and lower for smaller peptides. It's not uncommon for a 20 residue peptide to have an Xcorr of 5, while a 6 residue peptide might be around 1.5.

2.3.9.5 Characterisation and Cataloguing of Identified Proteins

GI accession numbers or *P. falciparum* IDs were used to retrieve molecular function, cellular component and biological process annotations according to gene ontology (GO) terms using the PIR batch retrieval tool (http://pir.georgetown.edu/cgi-bin/batch.pl) by selecting a specific identifier or a combination of identifiers:

a. ***Selecting the database:*** Before entering the protein identifiers the most convenient database was selected. If retrieving information about protein families, the PIRSF database was selected, however when analyzing individual proteins, the iProClass database was selected.

1. ***Rules for entering IDs*:** Multiple Entry IDs should be separated by lines or spaces. The maximum numbers of entries allowed is 50. IDs may be specified as a single category or as mixed categories. However, if your entries have the same type of ID, it is recommended that you define the ID field to speed up the retrieval process. Retrieved results and/or their respective sequences were saved as a tab de-limited text file for further analysis. To check the correspondence between search identifiers and the ones from the selected database, the list was generated using the "Show Match List"

2. *GO Slims:* These are smaller versions of the Gene Ontologies containing a subset of the terms in the whole GO. They give a broad overview of the ontology content without the detail of the specific fine grained terms. GO slims are particularly useful for giving a summary of the results of GO annotation of a genome or proteome when broad classification of gene product function is required. You can view the GO slim terms for molecular function, cellular component, and biological process by selecting the "Show GO Slim" button in the analysis tool bar. You can then view statistics for the individual ontologies (function, component, and process) by checking entries of interest and selecting the ontology to show.

2.4 Mass Profiling of CM

2.4.1 Background

Application of MS techniques in a broad-based screening approach to biomarker discovery has the potential to detect biomarkers that would not otherwise be studied. Further, biomarkers discovered in plasma and CSF would allow measurement of biological responses in a relatively non-invasive manner. MALDI-ToF MS and Surface Enhanced Laser Detector and Ionisation (SELDI)-ToF MS allow for the elucidation of individual protein masses from complex mixtures. By measuring the ToF of each protein or peptide after excitation with a laser pulse, a profile can be generated containing peaks representing the proteins or peptides found in an individual's plasma or CSF. Applying biostatistical methods to the acquired spectra allows MALDI-ToF MS technology to be used to investigate biomarkers in complex mixtures. ToF technology has been applied to blood biomarker detection of cancers of the ovary (Petricoin et al., 2002c, Ye et al., 2003, Alexe et al., 2004), prostate (Adam et al., 2002, Banez et al., 2003, Petricoin et al., 2002b), pancreas (Rosty et al., 2002, Valerio et al., 2001), breast (Li et al., 2002b), renal cells (Won et al., 2003), and head and neck (Wadsworth et al., 2004), as well as biomarkers for tropical diseases such as sleeping sickness (Papadopoulos et al., 2004). Most of these

studies have been carried out using SELDI-ToF instruments. Initial studies reported near perfect and perfect classification of ovarian cancer from control using SELDI profiles (Petricoin et al., 2002c, Petricoin et al., 2002d). However, it is clear from recent re-analysis of these data that early optimism of these results may be premature. Analysis of the same data set by outside groups could not reproduce the results (Baggerly et al., 2004, Sorace and Zhan, 2003). Furthermore, the use of SELDI for diagnostic pattern development has been met with some skepticism as it relies on pattern differences, which cannot be validated by other methods and are not biologically defined (Diamandis, 2003, Diamandis, 2004a, Diamandis, 2006). Alternatively, some SELDI studies are beginning to identify the protein(s) corresponding to a few classifying peaks (Rosty et al., 2002, Ye et al., 2003, Nomura et al., 2004). These identifications may strengthen the case for using ToF-MS in biomarker discovery. Despite the criticism and setbacks, ToF-MS has the potential for enormous impact on the field of biomarker discovery as it can offer a high-throughput analysis method applicable to complex biological mixtures. The overall pattern of protein expression can be a useful diagnostic tool, and the peaks produced on the spectra are diagnostic without necessarily identifying the proteins to which they correspond.

The profiling study described in this thesis has utilised MALDI-ToF to create protein profiles in plasma and CSF of children with CM. In brief the technique used involves (a) sample preparation (b) MALDI –ToF in linear mode (c) statistical analysis of peaks of interest. These are described in detail below

2.4.2 Sample Preparation

Sample preparation is important for the overall performance of mass profiling as it is essential to isolate the proteins while minimising protein disaggregation. Ideally one would like to mix the sample with matrix and apply sample on a MALDI target but this yields no results as the proteins in the sample need to be isolated. For this purpose,

detergents possessing the following properties are very useful: (1) high solubilising power; (2) no denaturation of proteins; (3) high critical micelle concentration; (4) non-ionic; (5) optical transparency; (6) chemical purity; (7) high solubility in water; (8)stability; (9) no effect on other assays (for example, protein assay); and (10) easily obtainable (Saito and Tsuchiya, 1984). The detergent n-octyl_-D-glucopyranoside possesses most of the desirable properties described (Baron and Thompson, 1975, Racker et al., 1979).

The matrix of choice for protein studies is sinapinic acid which is suitable for proteins with a mass greater than 10000 kDa. Sinapinic acid facilitates the resolution of the protonated molecule ion signal from the photochemically generated adduct signal by using a simple, linear time-of-flight mass spectrometer (Beavis and Chait, 1990).

2.4.3 Statistical Analysis of MALDI-ToF Spectra

Analysis of MALDI data can be separated into two distinct steps: (1) pre-processing and (2) processing or statistical analysis. The process of pre-processing is to reduce experimental variance within a data set, conditioning it for subsequent statistical analysis. Raw spectra are conditioned through removal of background, normalisation of intensity and alignment. The goal of statistical analysis of MALDI spectra is to use established statistical methods to either make a biomarker discovery or to classify samples into groups.

2.4.4 Experimental Protocols

2.4.4.1 Sample Preparation

10μg of plasma or CSF protein was added to 5ul of 1% n-Octyl β –D- glucopyronaside (NOG) and topped up to 100ul with dH_2O. 2ul of the sample mixed with 2ul of matrix solution made of 50% saturated solution of sinapinic acid prepared in 30% Acetonitrile in 0.1% TFA was then spotted on a MALDI-ToF target plate for and air dried.

2.4.4.2 MALDI-ToF Linear mode analysis

Mass spectrometry was performed using a AXIMA CFR Plus™ (Shimadzu Biotech-Kratos Analytical, Manchester, UK) mass spectrometer operated in a linear mode_2ghz. The following parameters were set for the data acquisition: mass range from 0 to approximately 180 000 d; laser power at 90; profile at 300; and five shots per spot. The instrument was calibrated using *m/z* ratios for the standards bovine serum albumin (66,429.09 Da). *m/z* ratios were determined for all data points. Pulsed extraction was set at 66,000Da. Average smoothing of spectra was applied using the company supplied algorithm with a peak width set at 10Da. The peak detection was set at Threshold-Apex at an offset of 0.1mV.

3 Proteomic Analysis of Plasma from Mice Infected With *Plasmodium Berghei* ANKA Strain Using 2-DE MALDI-ToF Mass Spectrometry

3.1 Introduction

The principle theme of this thesis is the use of proteomic platform technologies to provide insights into the biochemical changes that occur during severe malaria that would facilitate future development of novel diagnostic and techniques for monitoring and treatment of disease strategies. The genetic constitution of an individual may determine that individual's susceptibility to diseases and the time course of disease progression. Thus, the biochemical changes that occur in an individual in response to disease may be influenced by the individual's genetic background. Genes determine protein expression and proteome analysis is a direct measurement of proteins in terms of their presence and relative abundance. Proteomics is regarded as a complimentary technology to genome analysis. Proteins contain several facets that collectively indicate the actual functional state rather than the potential functional state as indicated in mRNA analysis. Moreover, although the pattern of gene activity may be perturbed in a tissue with a pathological lesions, there can often be a poor correlation between the level of transcription and translation i.e. the relative abundance of a protein within the specific cell type or tissue (Kennedy, 2001) . As a consequence of this, genomic data alone cannot be extrapolated to represent the functional status of the biological system. The overall aim of a proteomic study is the characterisation of the proteins in a system independent of but complementary to gene expression.

Characterisation of changes in protein levels in an individual with severe malaria can be complicated by a number of issues. There are the multiple factors involved and the condition itself of "severe malaria" probably reflects a number of distinct pathophysiological conditions all of which are compounded by the fact that there are rapid changes in disease condition. as it progresses. An animal model of malaria may provide be a useful tool in the initial efforts to characterise and understand the proteomic changes induced by the disease. . Cerebral malaria can be induced in susceptible strains of mice infected with the rodent malaria species *Plasmodium berghei* ANKA strain. These murine models of cerebral malaria have been used by many investigators to help understand the pathogenesis of the human condition (Rest, 1982, Reed et al., 1997, Thumwood et al., 1989, Jennings et al., 1997). As a prequel to the human studies described in chapters 4 and 5, this chapter describes a global investigation of the differences in protein expression profiles of plasma collected from infected and non-infected mice. This study had two functions firstly to develop and validate experimental procedures which we could then use to analyse human samples and secondly to start to understand the proteomic changes that accompany malaria infection in this model system.

3.2 Methodology

The general workflow for this proteomic study is summarized in **Fig. 3.1** and the background to these methodologies is described in detail in chapter 2 section 2.1. Briefly, a two platform strategy was employed based on (a) two-dimensional polyacrylamide gel electrophoresis (2-DE) to separate proteins coupled with gel imaging software to identify qualitative and semi-quantitative alterations in protein expression patterns followed by (b) mass spectrometry, to formally characterise the protein repertoire deemed to be of relevance to the biological processes active during murine CM. The high resolution 2-DE (Herbert, 1999, Gygi et al., 2000a) using immobilized *p*H gradients (IPG-Dalt) enabled the

resolution and array of *Rodent* and *Plasmodial* proteins from plasma collected from infected and non-infected mice. Subsequent analysis of the visualised protein expression patterns was performed using 2DE image analysis software.

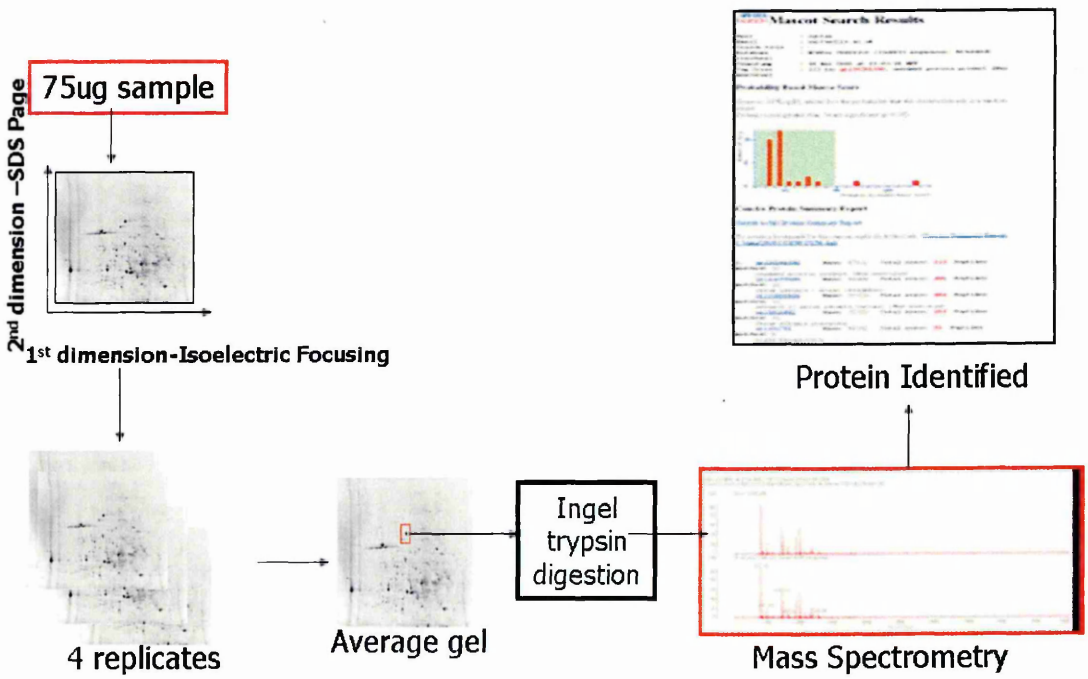

Figure 3.1 The 2-DE-MALDI Workflow. Proteins from the sample of interest are separated in the 1st dimension according to their pI and in the 2nd dimension according to there molecular weight. Replicate gels of each sample are then analysed using 2DE gel analysing software and average gels maps of each sample are then analysed for differential expression. Spots of interest are digested and applied onto a mass spectrometer. Resulting mass is then searched against sequence databases and a protein identity is made.

3.2.1 Preparation of plasma samples from mice

Plasma was collected from 4-week-old CD1 male mice (n= 3) infected with *Plasmodium Berghei* (0.1 ml of a culture with 2% parasitaemia). A control set of plasma was collected from mice (n=2) without infection. The sample was separated into two aliquots and albumin was depleted from one aliquot using the Montage Albumin kit (Millipore, USA).

3.2.2 P rotein Determination Using the Bradford Assay

Total plasma protein concentration was determined as described in chapter 2 section 2.1.8.1

3.2.3 Tw o Dimensional Gel Electrophoresis

Following determination of the protein concentration (chapter 2 section 2.1.8.1), proteins were separated using 2-DE as described in detail in chapter 2 section 2.1.8.2-2.1.8.4. Analytical gels used to provide master maps were loaded with 75µg of protein and preparative gels to be stained with Coomassie stain for eventual MALDI analysis were loaded with 300µg of plasma protein.

3.2.4 P rotein Visualisation and Image Analysis

Analytical gels were silver stained as described in chapter 2, section 2.1.8.5 and preparative gels stained with Coomassie as described in section 2.1.8.5. The stained gels were scanned using a GS-710 Imaging Densitometer (BioRad, Hemel Hempstead, Hertfordshire UK). The gels were analysed using PDQuest® software version 6.2.1 (Bio Rad, Hemel Hempstead, Hertfordshire UK) an d Progenesis PG 220® V2006 so ftware (Nonlinear Dynamics, Newcastle upon Tyne, UK) as described in chapter 2 section 2.1.8.5. PDQuest® enabled the matching of identical spots in serial gels and normalisation of gels to compensate for non-expression related variations in protein spot intensity. The software also enabled the tracking and reporting of all the protein patterns in the samples. By using the comparison tool for master gels, the software enabled the comparison of

differences between disease and control groups. Semi-quantitative analysis of spots was performed using Progenesis PG220®.

3.2.5 Trypti c In-gel Digestion

Gels spots were digested as described in chapter 2 section 2.2.3.1

3.2.6 Characteri zation of protein spots by MALDI-ToF-MS

Spot digests were analysed using MALDI-ToF as described in chapter 2 section 2.2.3.1.

3.2.7. Database Searching and Protein Characterisation and Cataloguing

All searches were performed using the Mascot™ algorithm as described in chapter 2 section 2.3.9.3. Mass lists generated from MALDI were used for searches against the NCBInr database. All searches were performed with the search parameters as : Database: NCBInr, Taxonomy: Rodentia and Alveolata, Enzyme: Trypsin, Tolerance: ±0.2Da, Fixed Modifications: Carbamidomethyl (C), Variable Modifications: Oxidation of methionine, Charge State MH+, Maximum missed cleavages: 1. Protein categorisation cataloguing was done according to the GO terms using the tool provided by PIR as described in detail in chapter 2 section 2.3.9.5

3.3 Results

3.3.1 P rotein Separation

Plasma samples from infected mice together with control samples were applied to 2-DE and proteins visualised by silver stain for analytical gels and coomassie blue stain for preparative gels. **Figure 3.2** shows representative gel maps for plasma from infected mice, and control mice before and after albumin depletion as created using PDQuest®. The gel maps are synthetic gels prepared by analyzing four replicate silver stained gels of each

sample. Spots from the match-set are represented on the master gel. As an example, **Figure 3.3** shows the match-set used to create the control with albumin depleted gel map. **Figure 3.4** is an example of a scattergram which demonstrates the good correlation of protein spot locations between the match-set replicate gels. Master gels of controls, infected, albumin depleted control and infected had 752, 639, 921, and 610 spots respectively. Using the PDQuest® comparison tool, composite gels showing differences in the infected master gel map compared to the control master gel map before and after albumin depletion were made as shown on **Figures 3.5 and 3.6**. There were protein spots that were differentially expressed in the infected gel and some proteins were missing completely. Removing albumin improved the quality of gels and helped identify some additional spots of interest.

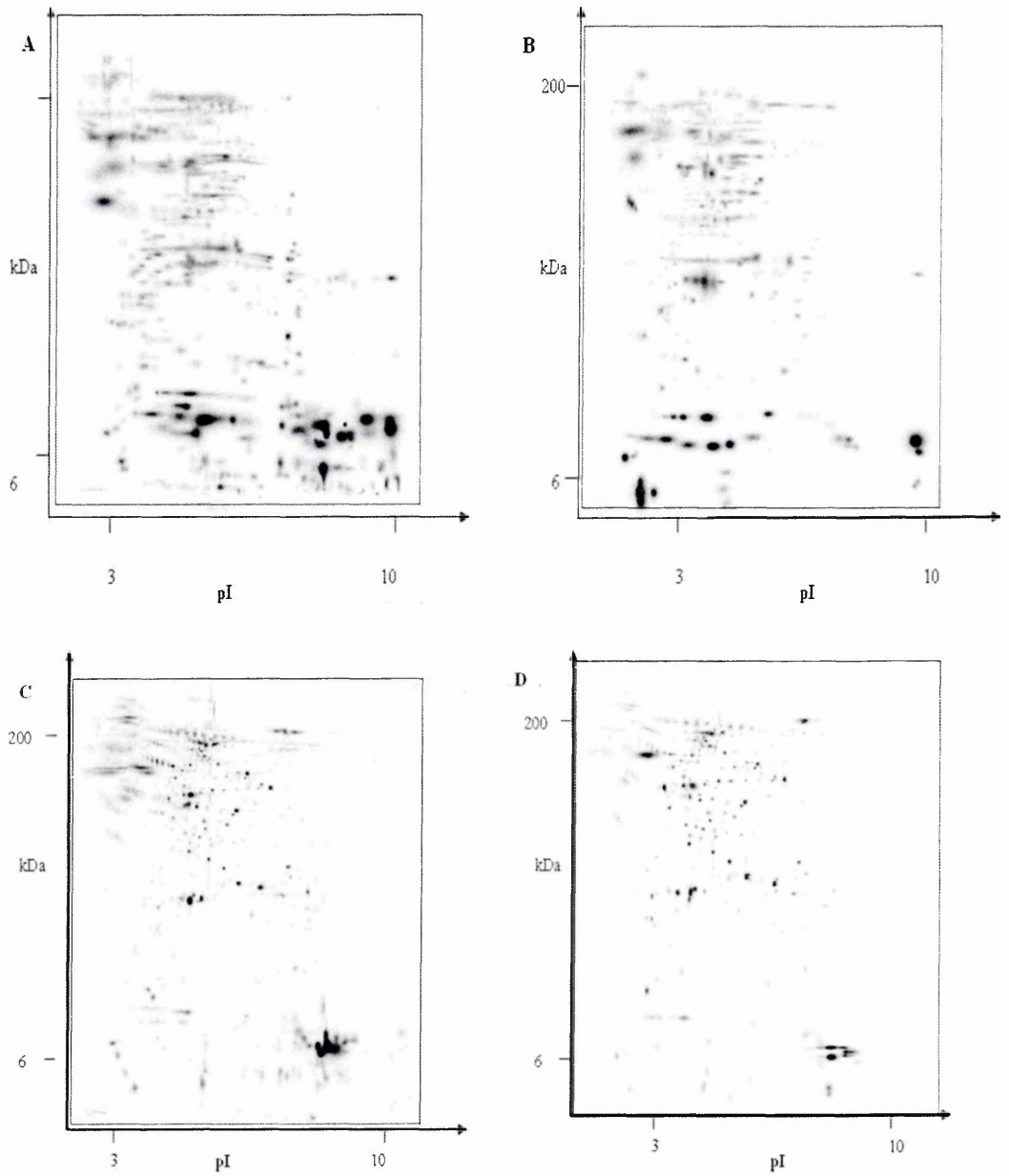

Figure 3.2 Representative gels maps created using PDQuest® for (A), Plasma from control mice (A), plasma from infected mice (B), albumin depleted plasma from control mice (C) and albumin depleted plasma from infected mice (D). The gel maps are synthetic images generated in silico from four replicate gels of a match-set. The plasma was drawn from 4-week old male mice infected with P. *Berghei* and had a parasitaemia of over 30%. Control plasma was from non-infected mice. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SDS-PAGE (12.5% T constant).

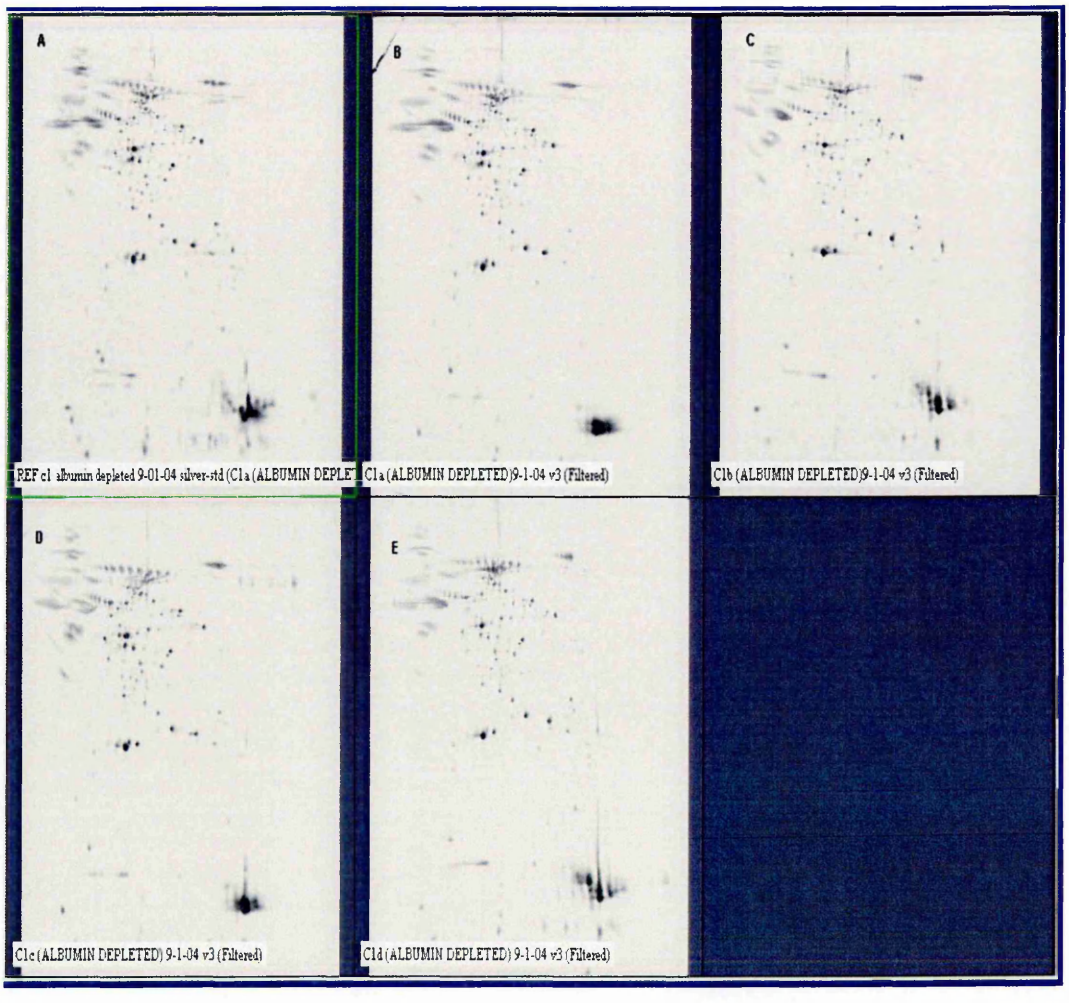

Figure 3.3 A gel match set created using PDQuest® for plasma with albumin depleted drawn from the control mice. Gel A represents the gel standard (gel map/average gel) and gels B-E represent the individual member gels. Spots were included in the gel map if found in every gel replicate. The 4 members were processed and run under standardized conditions

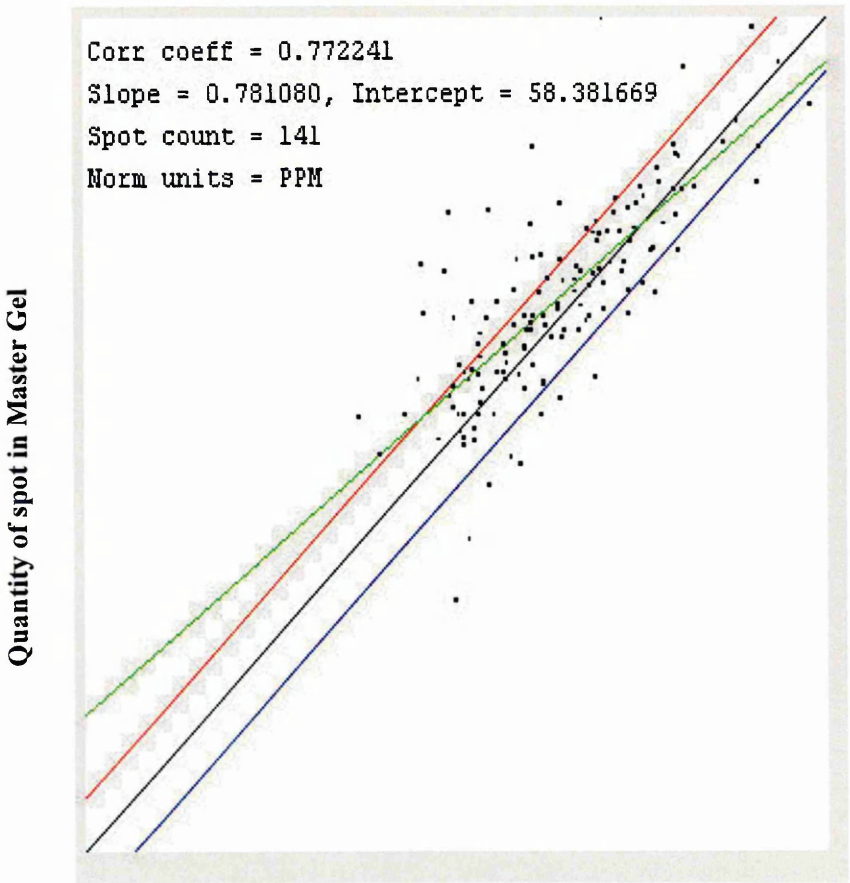

Figure 3.4 Shows a scattergram demonstrating the extent of correlation of protein spots between the Gel C and the Master Gel included in the Control with albumin depleted match-set (Figure 3.3). The quantity of each spot in the first gel (X-axis) is plotted on a log scale against its quantity in the second gel (Y-axis). If the spot quantity is the same in both gels, its point will fall on the black center line in the graph (slope = 1.00). If the quantities are not the same, the point will fall above or below the center line. The regression line generated from the plot is shown in green. The correlation coefficient for the regression line is displayed on the graph. A coefficient of 1.00 indicates that the two gels are perfectly similar, while a low coefficient value (e.g., 0.40) indicates that the two gels are not very similar. The number of spots that appear in both gels is also listed on the graph. Spots that fall within the reda and blue line fall within a 2-fold range.

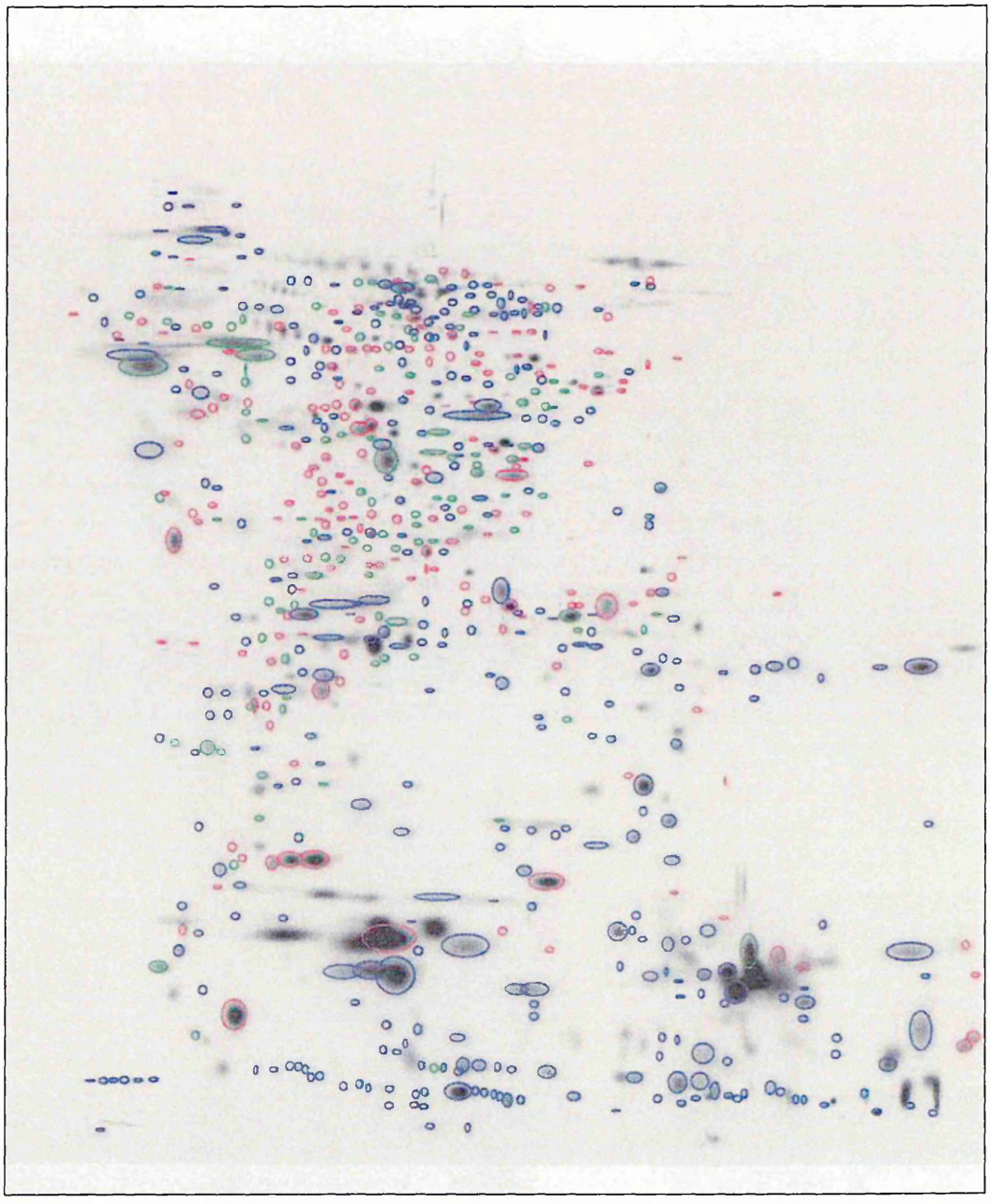

Figure 3.5 Composite gel showing the differences between gels from infected and control plasma. Composite gels were created using the comparison tool in PDQuest®. The composite gel demonstrates qualitative differences in the infected gel map compared to the control gel map. Red spots are spots found in the controls gel map and blue spots are those from the infected gel map. Green spots depict spots found in both

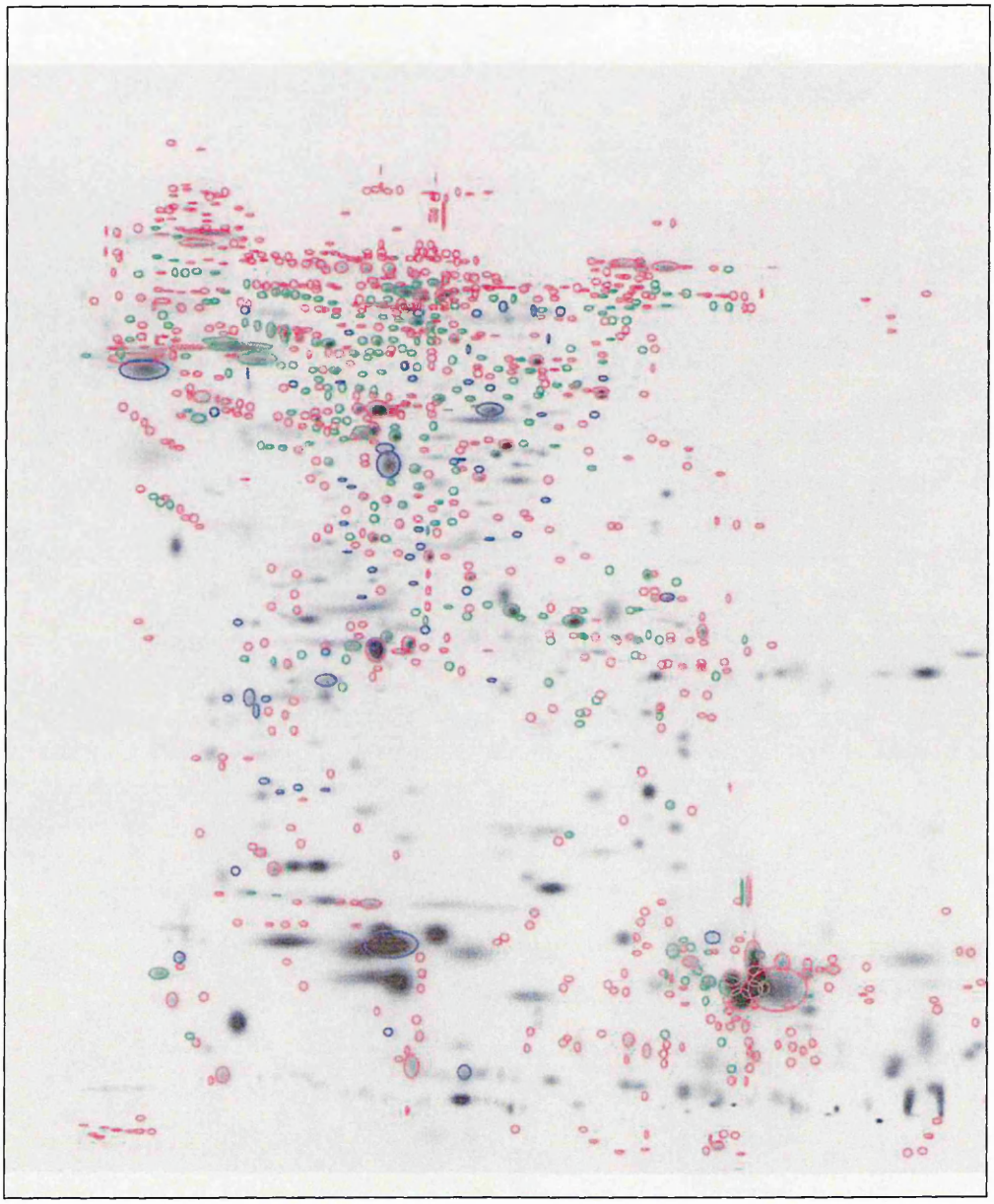

Figure 3.6 Composite gel showing the differences between albumin depleted infected and control plasma. Composite gels were created using the comparison tool in PDQuest®. The composite gel demonstrates qualitative differences in the infected (albumin depleted) gel map compared to the control gel map. Red spots are spots found in the control (albumin depleted) gel map and blue spots are those from the infected (albumin depleted) gel map. Green spots depict spots found in both

3.3.2 P rotein Identification by Mass Spectrometry (MS)

When stained with Coomassie between 60 and 104 spots per gel were visualised. (**Figures 3.7-3.10**). show the spots excised from the Coomassie stained gels. These spots were digested and prepared for MALDI analysis. It was possible to definitively identify 51, 63, 30 and 18 proteins from the infected, control, albumin depleted infected and albumin depleted control plasma samples respectively. **Table 3.1-4** gives a complete list of the proteins identified. In accordance with the cut-score provided by Mascot® described in detail in chapter 2 (section 2.3.9.3) the protein was considered to be tentatively identified if a significant score was achieved. For *Rodents* a protein score of 63 was considered to be significant and for *P. falciparum* a score of 57 was considered significant. Removing albumin improved the quality of gels and helped identify some additional proteins, but caused a reduction in our ability to identify spots by MALDI because of sensitivity issues. For semi-quantitative differential analysis silver stained replicate gels were analysed using Progenesis PG 220® 2DE software. The software created representative gels of replicates of each sample using one of the replicates as a base gel. Spots were given a match number if matched to 3 out of 4 replicates. Spots matched to each representative gel were then analysed for any changes in intensity. **Figure 3.11** shows some spots that were identified as having a greater than two fold increase in normalized spot intensity. **Table 3.2** gives a list of protein identities for these spots. **Figure 3.12** shows the differential expression of one of the spots in **Figure 3.11** and the resulting protein identification from Mascot®.

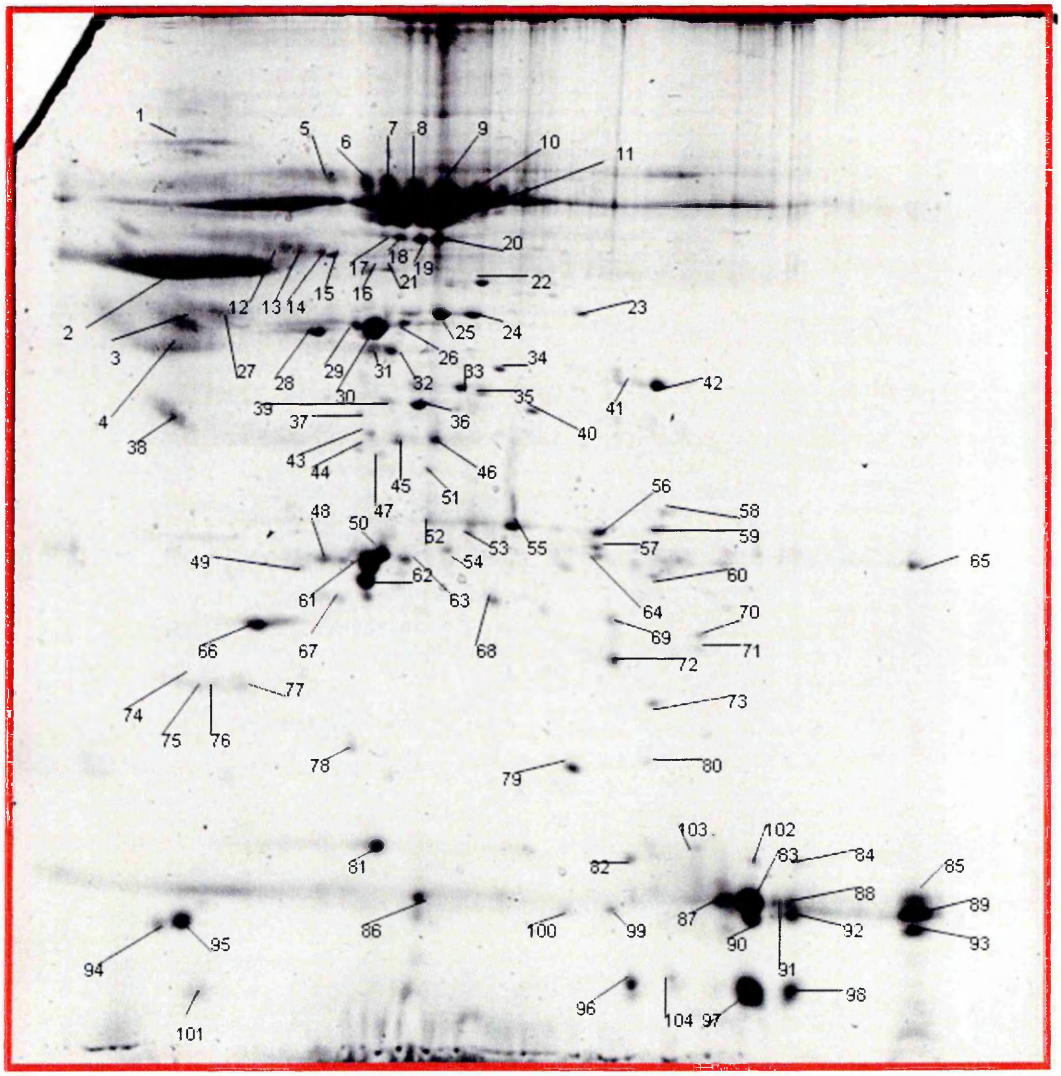

Figure 3.7 2-D PAGE gel of plasma from an infected mouse after loading with 300µg of sample protein. The plasma was drawn from 4-week old male mice infected with P. *Berghei* that had a parasitaemia of over 30%. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SDS-PAGE (12.5% T constant). The gel was stained with Coomassie blue stain. The numbered spot (1-103) were characterised by MALDI-MS and spot protein identities are listed in **table 3.1**

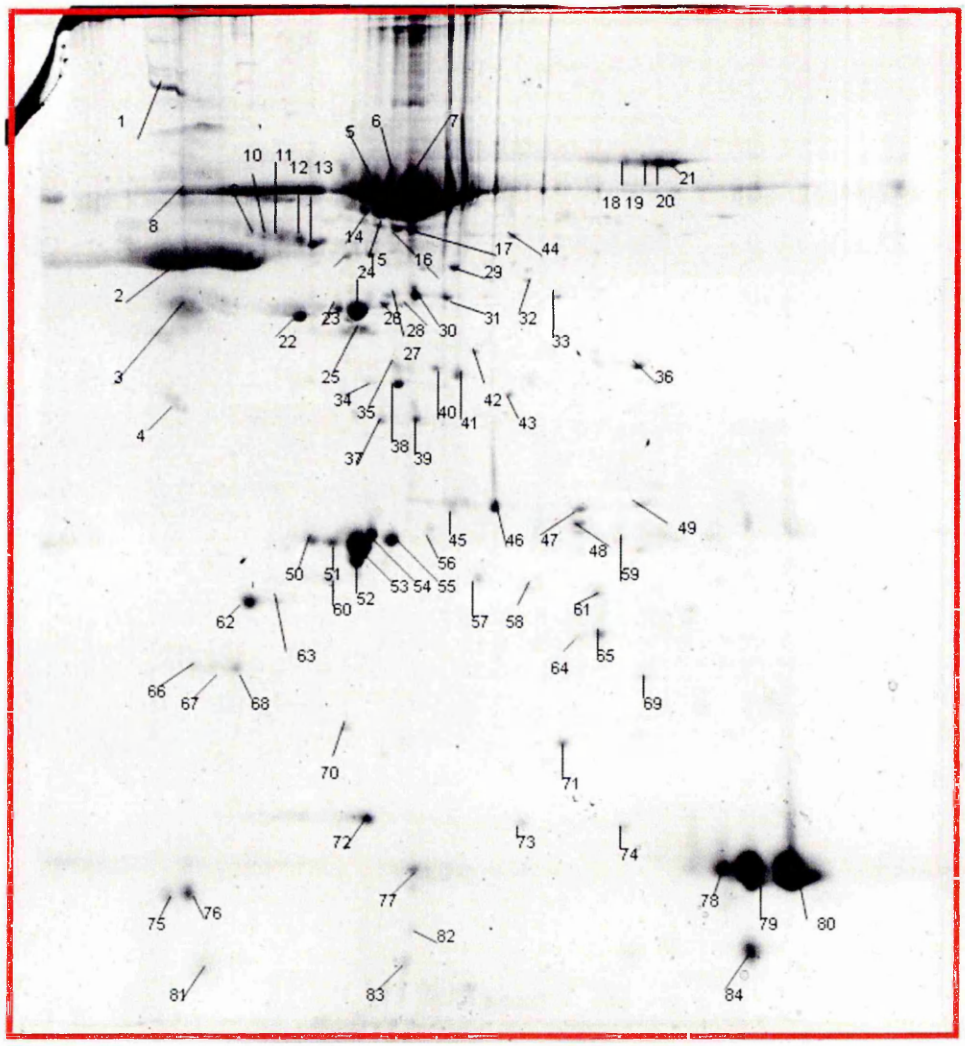

Figure 3.8 2-D PAGE gel of plasma from control mice after loading with 300µg of sample protein. The plasma was drawn from 4-week old male mice. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SDS-PAGE (12.5% T constant). The gel was stained with Coomassie blue stain. The numbered spot (1-84) were characterised by MALDI-MS and spot protein identities are listed on **table 3.2**

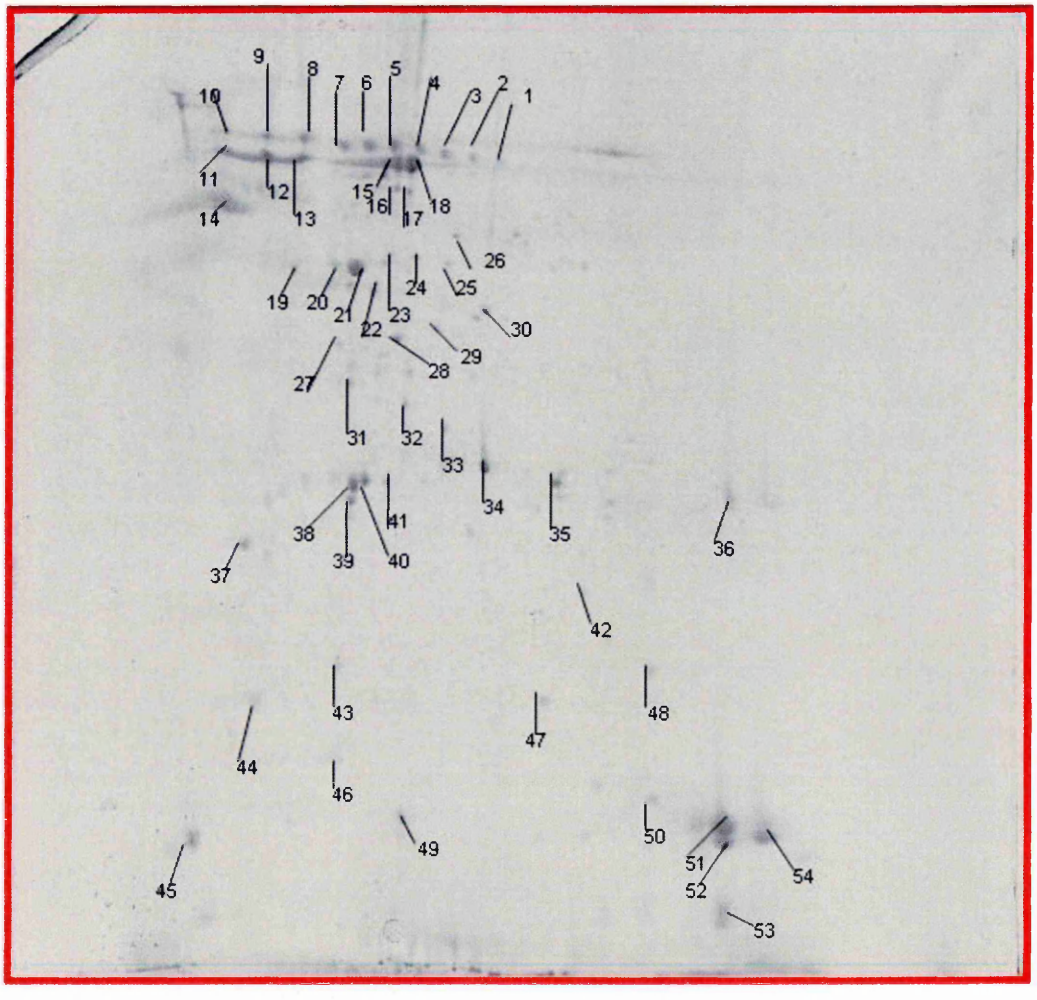

Figure 3.9 2-D PAGE gel of albumin depleted plasma from infected mice after loading with 300µg of sample protein. The plasma was drawn from 4-week old male mice infected with *P. Berghei* and had a parasitaemia of over 30%. Prior to 2DE, albumin was depleted using a commercial albumin depletion kit. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SDS-PAGE (12.5% T constant). The gel was stained with Coomassie blue stain. The numbered spot (1-54) were characterised by MALDI-MS and are listed on **table 3.3**

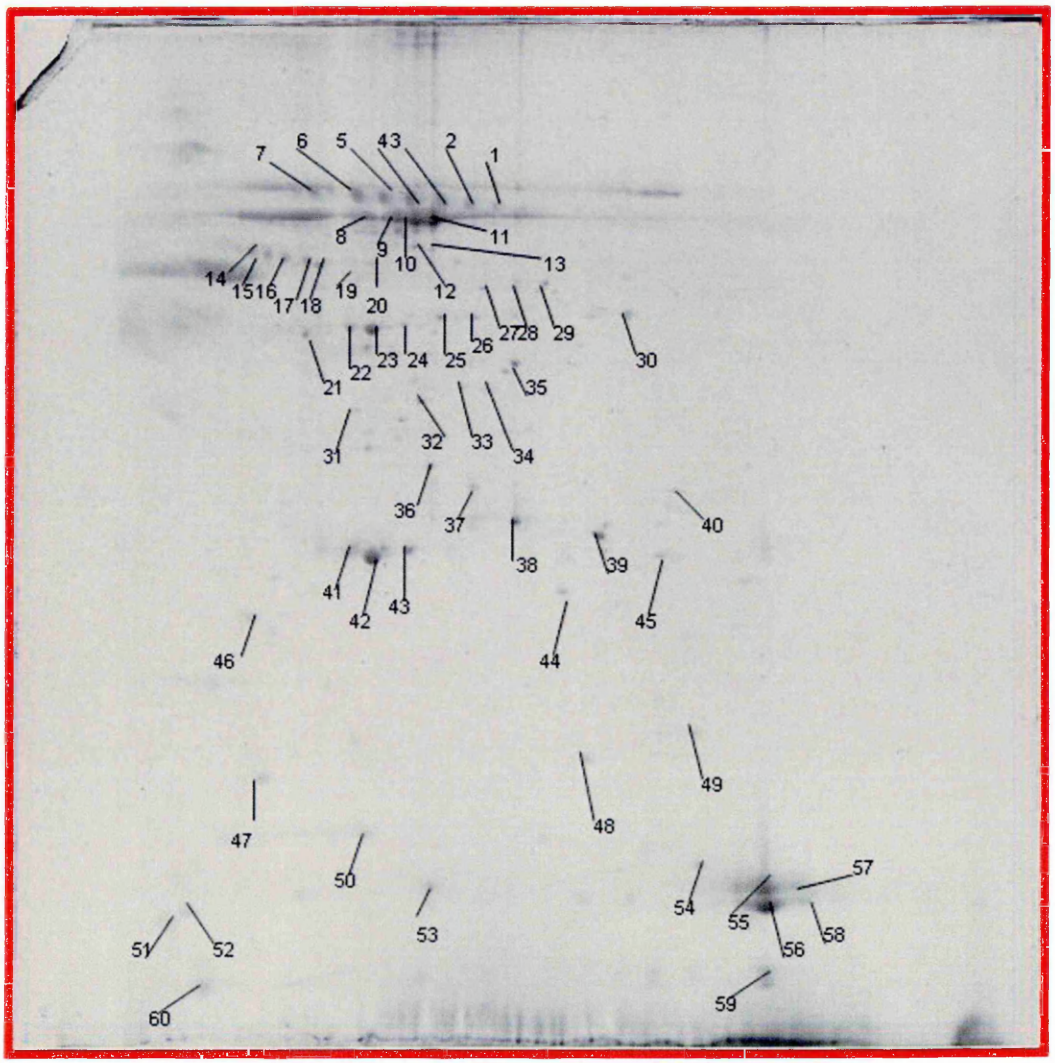

Figure 3.10 2-D PAGE gel of albumin depleted plasma from control mice after loading with 300µg of sample protein. The plasma was drawn from 4-week old male mice. Prior to 2DE, albumin was depleted using a commercial albumin depletion kit. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SDs-PAGE (12.5% T constant). The gel was stained with Coomassie blue stain. The numbered spot (1-60) were characterised by MALDI-MS and are listed on **table 3.4**

Table 3.1 List of proteins identified from the Coomassie stained gel of plasma from infected mice. The MALDI spectra were searched against the NCBInr database using the Mascot® search algorithm. Only proteins with a significant score as defined by Mascot® were included in the list.

Spot No.	Description	Accession	Score	Nominal Mass(Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
1	serine (or cysteine) proteinase inhibitor, clade A, member 1a; serine protease	6678079	83	46145	5.44	8	26
1	serine (or cysteine) proteinase inhibitor, clade A, member 1a; serine protease	6678083	83	45996	5.32	8	26
1	serine (or cysteine) proteinase inhibitor, clade A, member 1a; serine protease [*Mus Musculus*]	15029662	83	46109	5.32	8	26
1	similar to serine protease inhibitor 1-1 [*Mus Musculus*]	29612513	82	47208	5.39	8	25
1	serpina1a protein [*Mus Musculus*]	15029602	82	47299	5.20	8	25
1	serpina1a protein [*Mus Musculus*]	15929675	71	45764	5.31	7	24
1	serpina1a protein [*Mus Musculus*]	14602605	70	46094	5.32	7	24
3	hypothetical protein [*Plasmodium Falciparum* 3D7]	23612616	65	36741	8.79	7	16
4	hypothetical protein [*Plasmodium Falciparum*]	23612848	72	19694	8.23	6	31
12, 14	similar to macrophage galactose-type C-type lectin 2 [*Mus Musculus*]	38089618	78, 68	22892	5.77	7, 6	21, 20
13	hypothetical protein [*Plasmodium Yoelii Yoelii*]	23482237	58	30790	8.17	6	19
36	similar to retrovirus-related POL polyprotein [*Mus Musculus*]	38081301	77	47646	9.51	11	13
36	similar to spectrin alpha chain, brain (spectrin, non-erythrid alpha chain)	38074605	71	71650		10	
36	similar to ribosomal protein L37 [*Rattus Norvegicus*]	27680290	64	10678	11.29	5	43
36	similar to hypothetical protein ORF-1137 [*Mus Musculus*]	38080223	63	66035	9.65	10	9
42	60S ribosomal protein L35a	3914537	70	12646	10.90	5	31
42	ribosomal protein L35a [*Rattus Norvegicus*]	10863991	70	12660	10.90	5	31
42	similar to ribosomal protein L35a [*Mus Musculus*]	38084044	70	12780	10.63	5	31
42	unamed protein product [*Mus Musculus*]	26338025	70	12669	10.90	5	31
42	similar to ribosomal protein L35a [*Mus Musculus*]	38075333	66	14975	10.89	5	26
48, 50, 56, 72, 94, 95	apolipoprotein A-1 precursor - mouse	109571	137, 102,87,72,64,65	30358	5.52	11, 9, 8, 7, 7, 6	34, 29, 27, 23, 19, 19

Table 3.1 continued

Spot No.	Description	Accession	Score	Nominal Mass(Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
48, 50, 56, 67, 72, 94, 95	apolipoprotein A-1 [*Mus Musculus*]	6753096	136, 102, 87, 76, 72, 63, 65	30569	5.64	11, 9, 8, 8, 7, 7, 6	34, 29, 27, 28, 23, 19, 19
48, 50, 56, 67, 72, 94, 95	unamed protein product [*Mus Musculus*]	26345182	136, 102, 87, 76, 72, 64, 65	30597	5.51	11, 9, 8, 8, 7, 7, 6	34, 29, 27, 28, 23, 19, 19
48, 50, 56, 94, 95	apolipoprotein A-1 [*Mus Musculus*]	2145135	119, 86, 72, 64, 65	30498	5.64	10, 8, 7, 7, 6	29, 24, 22, 19, 19
48, 50, 56, 94, 95	apolipoprotein A-1 [*Mus Musculus*]	2145139	119, 86, 72, 64, 65	30526	5.51	10, 8, 7, 7, 6	29, 24, 22, 9, 19
48, 94	apolipoprotein A1 homolog [mus sp.]	1245804	69, 66	9577	4.68	5, 7	53, 53
50	ATP synthase subunit [*P. yoelii yoelii*	23482200	60	26084	9.03	6	20
55	unamed protein product [*Mus Musculus*]	26341396	80	67013	5.49	9	20
55	albumin 1; serum albumin variant [*Mus Musculus*]	33859506	78	70730	5.75	9	19
55	serum albumin precursor	5915682	78	70700	5.75	9	19
56	merozoite surface protein 3 [*Plasmodium Falciparum*]	2281613	65	20477	5.18	6	25
62	protein kinase c inhibitor-like protein, putative [*Plasmodium Falciparum* 3D7]	23612841	57	14748	5.95	5	25
65	similar to 60S Ribsomal protein L29 (P23) [*Rattus Norvegicus*]	34866986	67	17842	11.19	6	28
66	RNA binding motif protein , X chromosome [*Mus Musculus*]	6755296	72	42275	10.06	8	16
66, 103	heterogenous nuclear ribonucleoprotein G [*Mus Musculus*]	5579009	68, 67	31395	9.84	7, 7	20, 15
66	unamed protein product [*Mus Musculus*]	26339834	66	54509	9.36	9	10
66	Pfj1 [*Plasmodium Falciparum*]	2351192	60	71580	9.43	9	11
66	DNAJ protein [*P. falciparum* 3D7]	23510084	58	76894	9.33	9	10
73	hypothetical protein [*P. falciparum*]	23957743	77	12999	10.19	6	39
76	hypothetical protein [*P. yoelii yoelii*]	23482806	59	45217	9.26	7	13

Table 3.1 continued

Spot No.	Description	Accession	Score	Nominal Mass(Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
79	hypothetical protein [*P. yoelii yoelii*]	23482975	63	4350	7.93	4	86
80	similar to 60S Ribsomal protein L29 (P23) [*Mus Musculus*]	38089196	78	20985	11.46	7	25
80	similar to 60S Ribsomal protein L29 (P23) [*Rattus Norvegicus*]	34881340	71	16730	11.03	7	30
81	unamed protein product [*Mus Musculus*]	26335691	69	27986	5.57	7	25
81	heparan sulfate 6-0-sulfotransferase 1 [*Mus Musculus*]	20845347	64	23360	11.52	6	28
82	similar to HRPAP20 short form [*Rattus Norvegicus*]	34867098	65	20146	8.98	6	20
99	similar to Rp17a protein [*Rattus Norvegicus*]	34868324	86	29016	10.63	9	16
99	hypothetical protein [*P. falciparum* 3D7]	23613213	66	35968	9.51	7	15
99	similar to hypothetical protein FLJ25333 [*Mus Musculus*]	38075385	65	43041	8.11	7	11
99	var -CD36 protein [P. falciparum]	3786199	63	22635	7.11	6	21
99	splicing factor, putative [*P. falciparum* 3D7]	23613195	60	31541	10.91	6	15

Table 3.2 List of proteins identified from the Coomassie stained gel of plasma from control mice. The MALDI spectra were searched against the NCBInr database using the Mascot® search algorithm. Only proteins with a significant score as defined by Mascot® were included in the list.

Spot No.	Description	Accession	Score	Nominal Mass(Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
2	AA591032 protein [*Mus Musculus*]	13543161	62	22960	9.4	6	22
7, 23, 24, 46, 70, 71	albumin 1; serum albumin variant [*Mus Musculus*]	33859506	104, 104, 92, 136, 77, 76	70730	5.75	11, 11, 11, 14, 9, 9	22, 22, 25, 26, 18, 17
7, 23, 24, 46, 70, 71	Serum albumin precursor	5915682	104, 104, 92, 136, 77, 76	70700	5.75	11, 11, 11, 14, 9, 9	22, 22, 25, 26, 18, 17
7, 23, 24, 70, 71	unnamed protein product [*Mus Musculus*]	26341396	92, 113, 101, 78, 78	67013	5.49	10, 11, 11, 9, 9	21, 23, 26, 19, 18
7, 23, 24, 46	serum albumin -mouse fragment	11277085	88, 106, 96, 75	52988	5.49	9, 10, 10, 9	26, 25, 29, 22
7, 23, 46	alpha-fetoprotein	191765	75, 92, 64	48792	5.47	8, 9, 8	22, 23, 20
22	similar to apolipoprotein A-IV [*M. musculus*]	29477189	80	42889	5.55	8	26
22	apolipoprotein A-IV precursor- mouse (strain 129)	109575	78	44545	5.48	8	25
22	Apoa4 protein [*M. musculus*]	14789706	78	45001	5.34	8	25
22	apolipoprotein A-IV precursor [Apo-AIV]	1703331	78	45001	5.41	8	25
22	Tpm3 protein [*Mus Musculus*]	20809750	71	33357	4.73	8	22
28	hypothetical protein XP_125606 [*Mus Musculus*]	20858591	65	18468	9.33	7	22
30	lectin, galactose binding , soluble 7 [*Mus Musclus*]	31543120	66	15250	6.37	6	36
35	similar to 60S ribosomal protein L7a (Sufeit locus protein 3) PLA-X polypeptide	34869618	64	16346	10.53	6	29
40	similar to polyadenlate-binding protein 4 (PABP 4) [*Rattus Novegicus*]	27690704	81	42233	9.33	11	16
44	hypothetical protein XP_218509 [*Rattus Norvegicus*]	34855811	72	79394	9.44	9	9
46	unamed product [*Mus Musculus*]	2614396	140	67013	5.49	14	26
50	unamed product [*Mus Musculus*]	26325292	66	28268	9.39	7	26
50, 53, 54, 55, 61, 62, 64, 65	unamed product [*Mus Musculus*]	26345182	64, 123, 139, 92, 153, 118, 63, 86	30597	5.51	7, 11, 12, 9, 12, 10, 6, 8	24, 35, 38, 29, 31, 35, 18, 29
51	Non0/p54nrb homolog [*Rattus Norvegicus*]	2674209	71	21661	6.78	6	23
51	stress-induced phosphoprotein 1 [*Mus Musculus*]	13277819	69	63156	6.40	9	9
51	stress-induced phosphoprotein 1; stress - inducible protein; IEF SSP 3521 [*Mus Musculus*]	14389431	69	63170	6.40	9	9
52	cytochrome c oxidase subunit VIIc [*Mus Musculus*]	6680991	70	7328	11.00	5	36
53, 54, 55, 61, 62, 64, 65	apolipoprotein A-1 precursor mouse	109571	123, 139, 92, 153, 119, 63, 87	30358	5.52	11, 12, 9, 12, 10, 6, 8	35, 38, 29, 31, 30, 18, 9

Table 3.2 continued

Spot No.	Description	Accession	Score	Nominal Mass(Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
53, 54, 55, 61, 62, 64, 65	apolipoprotein A-1 [*Mus Musculus*]	6753096	123, 139, 68, 153, 118, 63, 86	30569	5.64	11, 12, 9, 12, 10, 6, 8	35, 38, 29, 31, 30, 18, 29
53, 54, 55, 61, 62, 65	apolipoprotein A-1 [*Mus Musculus*]	2145135	107, 122, 77, 153, 101, 86	30498	5.64	10, 11, 8, 12, 9, 8	30, 33, 24, 31, 29
53, 54, 55, 61, 62, 65	apolipoprotein A-1 [*Mus Musculus*]	2145139	107, 122, 77, 153, 101, 86	30526	5.51	10, 11, 8, 12, 9, 8	30, 33, 24, 31, 31, 29
53, 54, 61, 62, 76	apolipoprotein A1homolog [mus sp.]	1245804	81, 63, 104, 86, 69	9577	4.68	6, 5, 7, 6, 5	70, 62, 54, 62, 43
59	similar to HRPAP20 short form [*Rattus Norvegicus*]	34879811	80	20165	8.67	7	24
62	similar to 60S ribosomal protein L7a (Sufeit locus protein 3) PLA-X polypeptide	38081992	66	35049		7	
64, 68	chain A, solution structure of a recombinant mouse major urinary protein	8569601	85, 142	18869	4.82	7, 10	36, 36
67, 68	major urinary protein [mice]	1839508	85, 163	18855	4.75	7, 11	36, 42
67, 68	major urinary protein complex with 2-(sec butyl)	494384	69, 162	19322	4.82	6, 11	28, 30
67, 68	major urinary protein 1 [*Mus Musculus*]	13654245	67, 158	20920	5.02	6, 11	28, 30
67, 68	similar to major urunary protein 6 precursor	20972274	67, 158	20921	4.89	6, 11	28, 30
67, 68	major urinary protein [*Mus Musculus*]	13276755	67, 138	20935	4.96	6, 10	28
67, 68	major urinary protein 2 precursor	127527	67, 138	20935	5.04	6, 10	28, 28
67	similar to ribosomal protein L37 [*Mus Musculus*]	20984795	65	11376	10.07	5	45
67, 68	unamed product [*Mus Musculus*]	12849307	63, 63	23710	5.05	6, 6	26, 26
67	similar To RIKEN Cdna 1700001e04 [*Mus Musculus*]	38076876	62	26084	6.56	6	21
68	major urinary proteins 11 and 8	127531	147	17720		10	
68	unamed product [*Mus Musculus*]	12851568	140	21033		10	
68	major urinary protein [*Mus Musculus*]	13276755	138	20935		10	30
68	major urinary protein [*Mus Musculus*]	295910	121	20965		9	
68	major urinary protein	199869	102	20707		8	
68	alpha-2u-globulin I precursor -mouse	90284	93	15984	4.89	7	33
68	MUP1 protein product [*Mus Musculus*]	22477743	86	10904		6	
68	unamed product [*Mus Musculus*]	12847522	74	18484		6	

Table 3.2 continued

Spot No.	Description	Accession	Score	Nominal Mass(Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
70	albumin [*Mus Musculus*]	26986064	80	24237	5.48	7	38
71	60S Ribosoml protein L4 (L1)	3914699	78	11101	10.78	6	39
71	similar to protein kinase [*Rattus Norvegicus*]	27672609	75	41327	9.19	8	15
71	similar to serine/threonine kinase 22b (spermiogenesis associated) testis specific serine [M. musuculus]	20346160	75	41452	9.12	8	15
72	transthyretin [*Mus Musculus*]	7305599	97	15880	5.77	7	20
72	unamed product [*Mus Musculus*]	12852317	64	15808	5.76	5	34
75	similar to hypothetical protein FLJ25333 [*Mus Musculus*]	38075385	68	43041	8.11	7	17
75	RNA binding motif protein, X chromosome [*Mus Musculus*]	6755296	64	42275	10.06	7	13
77, 78, 79	chain b, chimeric human mouse carbonmonoxy hemoglobin (human zeta2 mouse beta 2)	18655687	72, 75, 93	15664	7.27	6, 6, 7	41, 41, 41, 62
77, 78, 79	hemoglobin, beta adlt major chain; beta major globin [*Mus Musculus*]	31982300	72, 74, 93	15795	7.14	6, 6, 7	41, 41, 61
77, 78, 79	unamed protein product [*Mus Musculus*]	12846616	72, 74, 93	15825	7.14	6, 6, 7	41, 41, 61
78, 79	hemoglobin beta	229301	75, 93	15767	8.09	6, 7	66, 67
78, 79	unamed protein product [*Mus Musculus*]	12846921	74, 93	15810	7.14	6, 7	61, 61
79	unamed protein product [*Mus Musculus*]	12846961	95	14443	7.14	7	55
79	unamed protein product [*Mus Musculus*]	12852164	76	15738	7.14	6	55
89	heterogenous nuclear ribonucleoprotein G [*Mus Musculus*]	5579009	67	31395	9.84	7	18

Table 3.3 List of proteins identified from the Coomassie stained gel of albumin depleted plasma from infected mice. The MALDI spectra were searched against the NCBInr database using the Mascot® search algorithm. Only proteins with a significant score as defined by Mascot® were included in the list.

Spot No.	Description	Accession	Score	Nominal Mass (Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
5, 8	hemopexin	23956086	64, 83	52049	7.92	8, 10	20, 28
6	hemopexin	22022646	62	31483	7.64	4	18
8	hemopexin	1881768	72	51255	7.32	9	26
15, 16, 21, 22, 30, 29, 28	albumin 1; serum albumin variant [*Mus Musculus*]	33859506	259, 69, 81, 80, 90, 71, 72	70730	5.75	24, 7, 14, 18, 14, 10, 18	46, 13, 25, 24, 25, 18, 24
15, 16, 21, 22, 30, 29, 28	serum albumin precursor	5915682	259, 69, 81, 80, 90, 71, 72	70700	5.75	24, 7, 14, 15, 15, 7, 7	46, 13, 25, 19, 25, 18, 24
15, 21, 22, 30, 29, 28, 31	unamed protein product [*Mus Musculus*]	26341396	245, 84, 83, 93, 73, 75, 64	67013	5.49	23, 14, 18, 14, 16, 18, 17	46, 27, 25, 27, 19, 25, 20
15	serum albumin - mouse (fragement)	11277085	157	52988	5.49	16	43
15	alpha-fetoprotein	191765	130	48792	5.47	14	39
15	albumin [*Mus Musculus*]	26986064	99	24237	5.48	10	54
16	albumin [*Rattus Norvegicus*]	19705431	68	70670	6.09	7	13
19, 41, 39, 42	apolipoprotein A-IV precurosor- mouse	6753096	72, 65, 65, 82	45455	5.77	11, 11, 10, 19	34, 35, 33, 37
19	apolipoprotein A-IV	109575	69	49223	5.86	11	31
19	Apoa4 protein [*Mus Musculus*]	6680702	66	45001	5.34	10	31
41, 38, 39, 42	apolipoprotein A-1 precursor - mouse	109571	67, 64, 67, 84	30358	5.52	11, 12, 10, 19	35, 34, 33, 37
41, 38, 39, 42	unamed protein product [*Mus Musculus*]	26345182	66, 64, 66, 84	30597	5.51	11, 12,10,19	35
40	Ig heavy chain [*Mus Musculus*]	21927918	65	12321	9.30	7	54
35	baculoviral IAP repeat-containing 3; apoptosisi inhibitor 1[Mus]	6680696	62	68523	5.65	14	12
36	erythrocyte memebrane protein 1 [*P. falciparum*]	39842963	59	13594	8.42	9	47
42	apolipoprotein A-1 [*Mus Musculus*	2145139	67	30526	5.51	17	32
42	apolipoprotein A-1 [*Mus Musculus*	2145141	65	30498	5.64	17	32
47	hypothetical protein [p yoelii]	23483254	62	57448	9.26	10	20
47	unamed protein product [*Mus Musculus*]	12862141	62	14063	10.06	7	44
48	similar to ERIC1 [*Rattus Norvegicus*]	34878530	70	65152	4.63	12	20
46	similar to pantothenate kinase 2; PANK2 [*Rattus Norvegicus*]	34858736	73	60958	9.12	13	20
46	putative yir3 protein [P yoelii]	23483075	66	38263	8.70	13	22
49	similar to hypothetical protein FLJ25333 [*Rattus Norvegicus*]	34852795	71	60946	8.74	13	20

Table 3.3 continued

Spot No.	Description	Accession	Score	Nominal Mass (Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
50	brain stress early protein Gbi; rho GTPase activating protein 14; EST AI452337	27597098	66	111502	6.07	12	16
51	putative yir4 protein [P yoelii]	23482729	60	15653	5.15	11	30
54	erythrocyte memebrane protein 1 [*P. falciparum*]	40646494	60	15219	8.58	8	44
52	hypothetical protein [p yoelii yoelii]	23481917	58	36313	8.05	14	23

Table 3.4 List of proteins identified from the Coomassie stained gel of albumin depleted plasma from control mice. The MALDI spectra were searched against the NCBInr database using the Mascot® search algorithm. Only proteins with a significant score as defined by Mascot® were included in the list.

Spot No.	Description	Accession	Score	Nominal Mass (Mr)	Calculated pI value	Matched Peptides	Sequence coverage (%)
2, 5	Hemopaxin *Mus Musculus* (Mouse)	22022646	72, 88	31483	7.64	7, 9	27, 32
5	Hemopaxin *Mus Musculus* (Mouse)	23956086	95	52049	7.92	11	23
5	Hemopaxin *Mus Musculus* (Mouse)	1881768	83	51255	7.32	10	21
12, 13, 14, 30, 24, 38, 40	albumin 1; serum albumin variant [*Mus Musculus*]	33859506	100, 70, 63, 82, 71, 69, 93	70730	5.75	11, 9, 10, 11, 11, 11, 12	19, 12, 12, 22, 22, 17, 23
12, 13, 14, 30, 24, 40	serum albumin precursor	5915682	100, 70, 63, 82, 71, 69, 93	70700	5.75	11, 9, 10, 11, 10, 11, 12	19, 12, 12, 22, 22, 17, 23
12, 13, 14, 30, 24, 31, 40	unamed protein product [*Mus Musculus*]	26341396	90, 74, 66, 73, 73, 63, 95, 63	67013	5.49	10, 9, 10, 10, 9, 9, 12, 9	19, 13, 13, 22, 25, 20, 25, 20
12, 13	alpha-fetoprotein	191765	75, 72	48792	5.47	8, 8	22, 16
12, 13, 23, 24, 40	serum albumin - mouse (fragement)	11277085	72, 69, 70, 68, 67	52988	5.49	8, 8, 9, 9, 9	22, 15, 25, 22, 22
12	albumin [*Rattus Norvegicus*]	19705431	69	70670	6.09	9	15
38	albumin [*Mus Musculus*]	26986064	98	24237	5.48	10	47
44, 43, 42, 41	apolipoprotein A-1 precursor - mouse	109571	68, 70, 80, 70	30358	5.52	10, 10, 10, 9	32, 29, 32, 32
44, 43, 42, 41	unamed protein product [*Mus Musculus*]	26345182	68, 69, 80, 70	30597	5.51	10, 10, 10. 9	32, 29, 32, 31
44, 43, 42, 41	apolipoprotein A-1 [*Mus Musculus*]	6753096	66, 68, 79, 69	30569	5.64	10, 10, 10, 9	32, 29, 32, 31
42	apolipoprotein A-1 [*Mus Musculus*]	2145139	67	30526	5.51	10	27
42	apolipoprotein A-1 [*Mus Musculus*]	2145135	66	30498	5.64	9	27
55	similar to chaperonin containing TCP-1 gamma sub unit	34851497	70	59295	6.34	6	8
57	hemoglobin beta minor chain - mouse	70433	76	15851	7.97	7	45
57	hemoglobin adult minor hain; beta minor globin [*Mus Musculus*]	17647499	76	15982	7.85	7	45
57	hemoglobin beta- chain [*Mus Musculus*]	1183933	75	16438	8.56	7	44

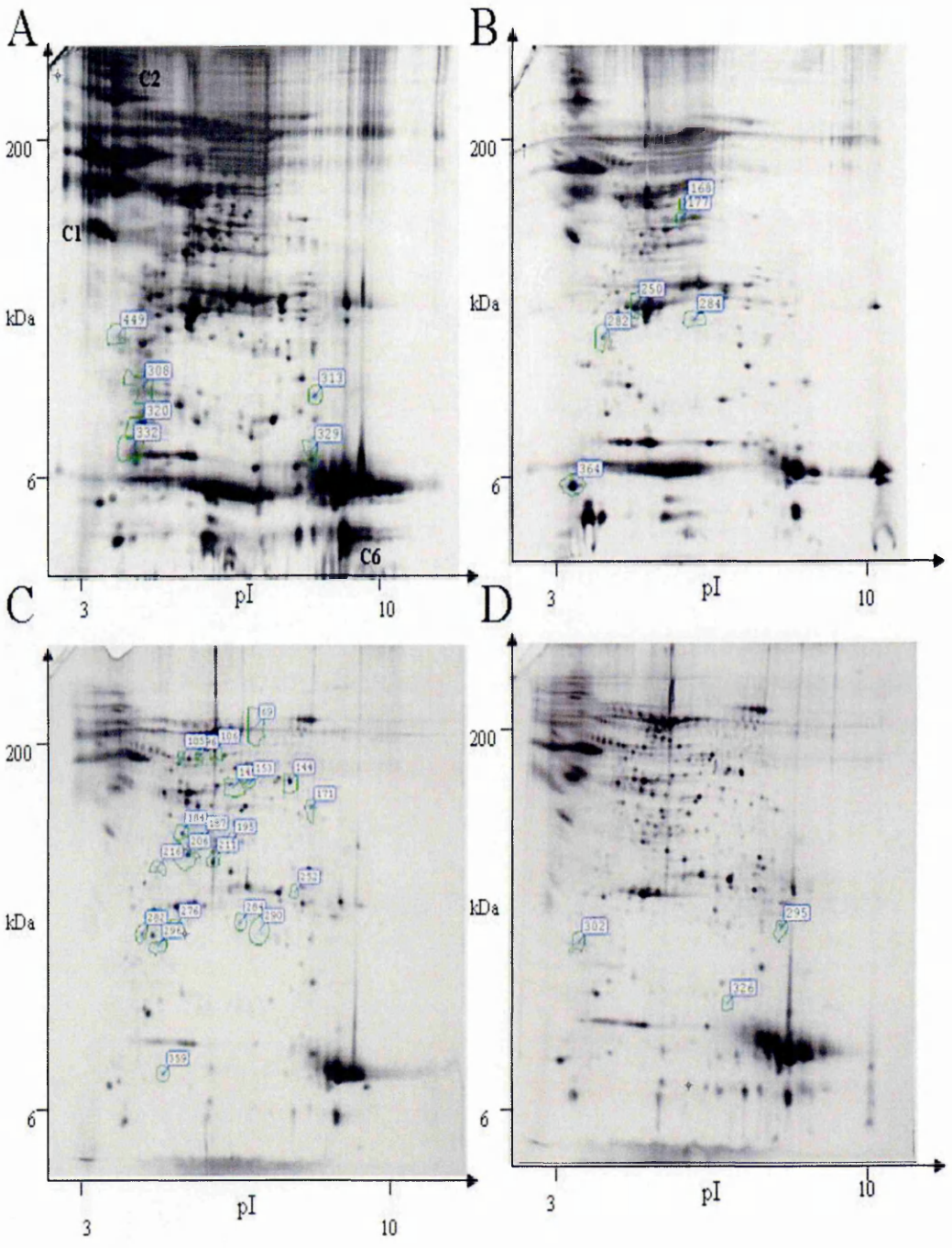

Figure 3.11 Differentially expressed spots of gels from plasma from infected mice(A) compared to control mice(B), albumin depleted plasma from infected mice (C) compared to albumin depleted plasma from control mice(D). The analysis was done using Progenesis PG220® which made reference gels of replicate gels and matched all spot in the replicates to the reference gels. The spots identified had an increase in normalized spot intensity by two-fold or more. Spots of interest were digested and analysed using MALDI followed by Mascot®. **Table 3.5** lists the proteins identified from the spots.

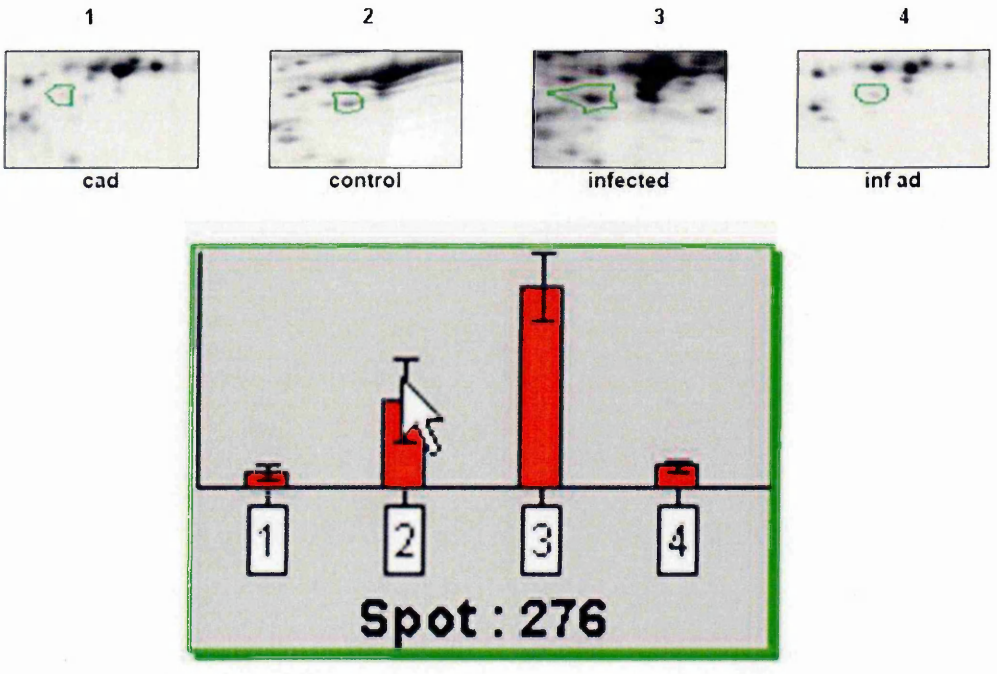

Figure 3.12 Spot 276 from Figure 3.11. Progenesis PG220® 2-DE gel analysis software showed a >2-fold increase in normalized spot intensity in the plasma from infected mice compared to the plasma from control mice. The histogram shows the intensity levels in each group. The spot was identified as Apolipoprotein A-1 and the MALDI spectra and subsequent Mascot® result are shown on **Figure 3.13**

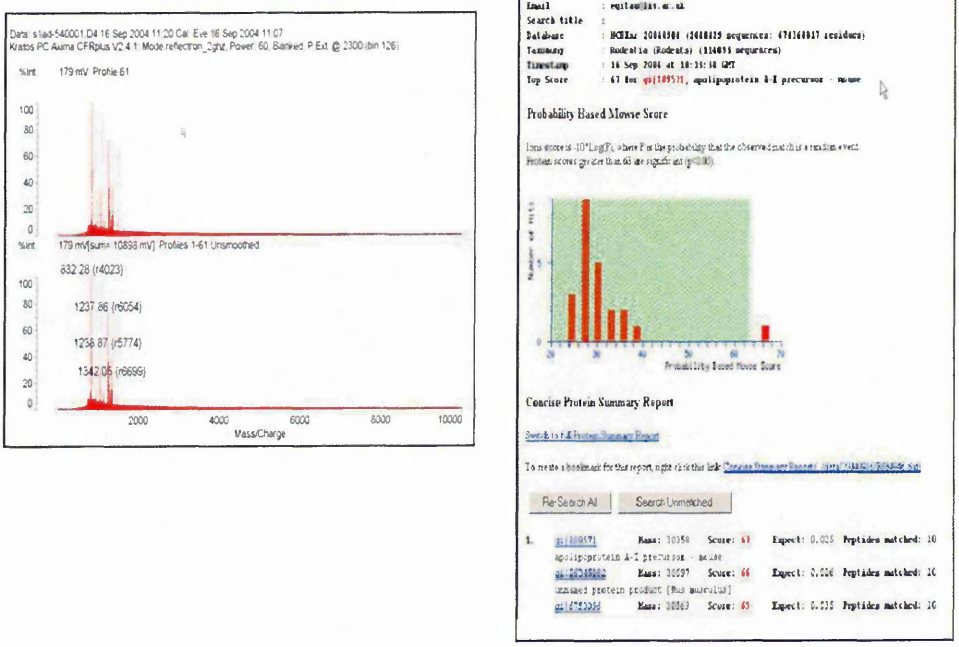

Figure 3.13 MALDI Spectra and Mascot Result for Spot 256 (Figure 3.12)

Table 3.5 Protein identification of spots shown to have a significant increase in expression in figure 3.11

Spot Number	Gel Type	Protein Identification	Accession number
177	Control	-	-
168	Control	similar to polyadenlate-binding protein 4 (PABP 4)	27690704
		Non0/p54nrb homolog [*Rattus Norvegicus*]	2674209
250	Control	stress-induced phosphoprotein 1 [*Mus Musculus*]	13277819
		stress-induced phosphoprotein 1; stress - inducible protein; IEF SSP 3521 [*Mus Musculus*]	14389431
284	Control	-	6753096
364	Control	Apolipoprotein A-1	6753096
295	Control (Albumin Depleted)	-	-
302	Control (Albumin Depleted)	unamed protein product [*Mus Musculus*]	26341396
326	Control (Albumin Depleted)	Serum albumin	11277085
449	Infected	-	-
308	Infected	-	-
320	Infected	-	-
332	Infected	-	-
313	Infected	-	-
329	Infected	-	-
69	Infected (albumin depleted)	-	-
106	Infected (albumin depleted)	albumin 1; serum albumin variant [*Mus Musculus*]	33859506
96	Infected (albumin depleted)	-	-
105	Infected (albumin depleted)	-	-
145	Infected (albumin depleted)	-	-
153	Infected (albumin depleted)	-	-
144	Infected (albumin depleted)	-	-
171	Infected (albumin depleted)	-	-
184	Infected (albumin depleted)	-	-
187	Infected (albumin depleted)	albumin 1; serum albumin variant [*Mus Musculus*]	33859506
195	Infected (albumin depleted)	albumin [*Mus Musculus*]	5915682
211	Infected (albumin depleted)	-	-
206	Infected (albumin depleted)	-	-
216	Infected (albumin depleted)	-	-
252	Infected (albumin depleted)	baculoviral IAP repeat-containing 3; apoptosisi inhibitor 1[*Mus Musculus*]	6680696
290	Infected (albumin depleted)	-	-
284	Infected (albumin depleted)	-	-
		apolipoprotein A-1[*Mus Musculus*	6753096
276	Infected (albumin depleted) and infected	apolipoprotein A-1 precursor - mouse	109571
296	Infected (albumin depleted)	-	-
282	Infected (albumin depleted)	-	-
359	Infected (albumin depleted)	similar to hypothetical protein FLJ25333 [*Rattus Norvegicus*]	34852795

3.3.3 Functio nal Cataloguing of Proteins

Protein functions were identified using the PIR id mapping tool and catalogued according to their gene ontology (GO) number. **Table 3.6** gives a summary of all the proteins identified on the gels. GO gave the molecular functions, cellular components and biological processes of the proteins. 153 proteins were identified from all the gels and of these 47 were not fully catalogued as they were either hypothetical proteins or had no GO matches. **Figure 3.14** show a graphic breakdown of the proteins according to their GO functional categories.

Table 3.6 Summary of Proteins Identified on Mouse 2-DE gels. The Gene Ontology (GO) numbers were derived from Protein Information Resource (PIR) -http://pir.georgetown.edu/ Batch retrieval tool. This tool converted the GI NCBInr accession numbers to GO numbers and then used GO slim to categorise the proteins. The table is colour coded to identify proteins found in only one group. Red= Plasma from infected mice, Green=plasma from control mice, Violet=albumin depleted plasma from infected mice, Blue=albumin depleted plasma from control mice

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
albumin [*Mus Musculus*]	26986064	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005737 : cytoplasm; 0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization		70	15	38
albumin [*Rattus Norvegicus*]	19705431	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005737 : cytoplasm; 0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0050896 : response to stimulus; 0006810 : transport; 0007154 : cell communication; 0044419 : interaction between organisms; 0008150 : biological_process; 0016043 : cell organization and biogenesis; 0051179 : localization			16	12
albumin 1; serum albumin variant [*Mus Musculus*]	33859506	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005737 : cytoplasm; 0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization	55	7, 23, 24, 46, 70, 71	15, 16, 21, 22, 30, 29, 28	12, 13, 14, 30, 24, 38, 40
alpha-fetoprotein	191765	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005737 : cytoplasm; 0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization		7, 23, 46	15	12, 13
apolipoprotein A-1 [*Mus Musculus*]	6753096	0005215 : transporter activity; 0005515 : protein binding; 0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006629 : lipid metabolism; 0006869 : lipid transport; 0006810 : transport; 0006066 : alcohol metabolism	48, 50, 56, 67, 72, 94, 95	53, 54, 55, 61, 62, 64, 65	19, 41, 39, 42	44, 43, 42, 41

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
apolipoprotein A-1 [*Mus Musculus*]	2145139	0005215 : transporter activity; 0005515 : protein binding; 0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006629 : lipid metabolism; 0006869 : lipid transport; 0006810 : transport; 0006066 : alcohol metabolism	48, 50, 56, 94, 95	53, 54, 55, 61, 62, 65	42	42
apolipoprotein A-1 [*Mus Musculus*]	2145135	0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0006869 : lipid transport	48, 50, 56, 94, 95	53, 54, 55, 61, 62, 65		42
apolipoprotein A-1 precursor - mouse	109571	-	-	-	48, 50, 56, 72, 94, 95	53, 54, 55, 61, 62, 64, 65	41, 38, 39, 42	44, 43, 42, 41
hemoglobin adult minor hain; beta minor globin [*Mus Musculus*]	17647499	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport				57
hemoglobin beta-chain [*Mus Musculus*]	1183933	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport				57
hemoglobin beta minor chain - mouse	70433	-	-	-				57
Hemopaxin *Mus Musculus* (Mouse)	22022646	-	-	-			6	2, 5
Hemopaxin *Mus Musculus* (Mouse)	23956086	0043167 : ion binding	-	0006810 : transport			5, 8	5
Hemopaxin *Mus Musculus* (Mouse)	1881768	0043167 : ion binding	-	0006810 : transport			8	5
serum albumin - mouse (fragement)	11277085	-	-	-		7, 23, 24, 46	15	12, 13, 23, 24, 40

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
serum albumin precursor	5915682	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005737 : cytoplasm; 0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization	55	7, 23, 24, 46, 70, 71	15, 16, 21, 22, 30, 29, 28	12, 13, 14, 30, 24, 40
similar to chaperonin containing TCP-1 gamma sub unit	34851497	-	-	-				55
unamed protein product [*Mus Musculus*]	26341396	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005737 : cytoplasm; 0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization	55	7, 23, 24, 70, 71	15, 21, 22, 30, 29, 28, 31	12, 13, 14, 30, 24, 31, 40
unamed protein product [*Mus Musculus*]	26345182	0005215 : transporter activity; 0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006869 : lipid transport	48, 50, 56, 67, 72, 94, 95	50, 53, 54, 55, 61, 62, 64, 65	41, 38, 39, 42	44, 43, 42, 41
AA591032 protein [*Mus Musculus*]	13543161	-	-	-		2		
similar to apolipoprotein A-IV [*M. musculus*]	29477189	0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006869 : lipid transport		22		
apolipoprotein A-IV precursor-mouse (strain 129)	109575	-	-	-		22	19	
Apoa4 protein [*M. musculus*]	14789706	0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006869 : lipid transport		22		

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
apolipoprotein A-IV precursor [Apo-AIV]	1703331	0005215 : transporter activity; 0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006869 : lipid transport; 0006810 : transport		22		
Tpm3 protein [*Mus Musculus*]	20809750		0005737 : cytoplasm; 0005856 : cytoskeleton	0050789 : regulation of biological process		22		
hypothetical protein XP_125606 [*Mus Musculus*]	20858591	-	-	-		28		
lectin, galactose binding , soluble 7 [*Mus Musclus*]	31543120	0030246 : carbohydrate binding				30		
similar to 60S ribosomal protein L7a (Sufeit locus protein 3) PLA-X polypeptide	34869618	-	-	-		35		
similar to polyadenlate-binding protein 4 (PABP 4) [*Rattus Novegicus*]	27690704	-	-	-		40		
hypothetical protein XP_218509 [*Rattus Norvegicus*]	34855811	-	-	-		44		
unamed product [*Mus Musculus*]	2614396	-	-	-		46		
unamed product [*Mus Musculus*]	26325292	-	-	-		50		
Non0/p54nrb homolog [*Rattus Norvegicus*]	2674209	-	-	-		51		
stress-induced phosphoprotein 1 [*Mus Musculus*]	13277819	0005488 : binding	0005634 : nucleus; 0005794 : Golgi apparatus	0050896 : response to stimulus		51		

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
stress-induced phosphoprotein 1; stress - inducible protein; IEF SSP 3521 [*Mus Musculus*]	14389431	0005488 : binding	0005634 : nucleus; 0005794 : Golgi apparatus	0050896 : response to stimulus		51		
cytochrome c oxidase subunit VIIc [*Mus Musculus*]	6680991	0005215 : transporter activity; 0016491 : oxidoreductase activity	0016020 : membrane; 0005739 : mitochondrion	0006091 : generation of precursor metabolites and energy		52		
apolipoprotein A1homolog [mus sp.]	1245804	0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0006869 : lipid transport	48, 94	53, 54, 61, 62, 76		
similar to HRPAP20 short form [*Rattus Norvegicus*]	34879811	-	-	-		59		
similar to 60S ribosomal protein L7a (Sufeit locus protein 3) PLA-X polypeptide	38081992	-	-	-		62		
chain A, solution structure of a recombinant mouse major urinary protein	8569601	-	-	-		64, 68		
major urinary protein [mice]	1839508	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		67, 68		
major urinary protein complex with 2-(sec butyl)	494384	-	-	-		67, 68		
major urinary protein 1 [*Mus Musculus*]	13654245	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		67, 68		

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
similar to major urunary protein 6 precursor	20972274	-	-	-		67, 68		
major urinary protein [*Mus Musculus*]	13276755	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		67, 68		
major urinary protein 2 precursor	127527	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		67, 68		
similar to ribosomal protein L37 [*Mus Musculus*]	20984795	-	-	-		67		
unamed product [*Mus Musculus*]	12849307		0005634 : nucleus			67, 68		
similar To RIKEN Cdna 1700001e04 [*Mus Musculus*]	38076876	-	-	-		67		
major urinary proteins 11 and 8	127531	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		68		
unamed product [*Mus Musculus*]	12851568	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		68		
major urinary protein [*Mus Musculus*]	295910	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		68		
major urinary protein	199869	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		68		
alpha-2u-globulin I precursor -mouse	90284	-	-	-		68		
MUP1 protein product [*Mus Musculus*]	22477743	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		68		
unamed product [*Mus Musculus*]	12847522	0005488 : binding; 0005215 : transporter activity; 0005549 : odorant binding		0006810 : transport		68		

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
60S Ribosoml protein L4 (L1)	3914699		0030529 : ribonucleoprotein complex			71		
similar to protein kinase [*Rattus Norvegicus*]	27672609	-	-	-		71		
similar to serine/threonine kinase 22b (spermiogenesis associated) testis specific serine [M. musuculus]	20346160	-	-	-		71		
transthyretin [*Mus Musculus*]	7305599	0005496 : steroid binding	0005576 : extracellular region	0006810 : transport		72		
unamed product [*Mus Musculus*]	12852317	0005496 : steroid binding	0005576 : extracellular region	0006810 : transport		72		
similar to hypothetical protein FLJ25333 [*Mus Musculus*]	38075385	-	-	-	99	75		
RNA binding motif protein, X chromosome [*Mus Musculus*]	6755296	0000166 : nucleotide binding; 0003676 : nucleic acid binding	0030529 : ribonucleoprotein complex		66	75		
chain b, chimeric human mouse carbonmonoxy hemoglobin (human zeta2 mouse beta 2)	18655687	-	-	-		77, 78, 79		
hemoglobin, beta adlt major chain; beta major globin [*Mus Musculus*]	31982300	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport		77, 78, 79		
unamed protein product [*Mus Musculus*]	12846616	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport		77, 78, 79		

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
hemoglobin beta	229301	-	-	-		78, 79		
unamed protein product [*Mus Musculus*]	12846921	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport		78, 79		
unamed protein product [*Mus Musculus*]	12846961	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport		79		
unamed protein product [*Mus Musculus*]	12852164	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport		79		
heterogenous nuclear ribonucleoprotein G [*Mus Musculus*]	5579009	0000166 : nucleotide binding; 0003676 : nucleic acid binding	0030529 : ribonucleoprotein complex		66, 103	89		
serine (or cysteine) proteinase inhibitor, clade A, member 1a; serine protease	6678079	0030234 : enzyme regulator activity		0050896 : response to stimulus	1			
serine (or cysteine) proteinase inhibitor, clade A, member 1a; serine protease	6678083	0030234 : enzyme regulator activity			1			
serine (or cysteine) proteinase inhibitor, clade A, member 1a; serine protease [musmusculus]	15029662	0030234 : enzyme regulator activity		0050896 : response to stimulus	1			
similar to serine protease inhibitor 1-1 [*Mus Musculus*]	29612513	0030234 : enzyme regulator activity		0050896 : response to stimulus	1			

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
serpina1a protein [*Mus Musculus*]	15029602	0030234 : enzyme regulator activity			1			
serpina1a protein [*Mus Musculus*]	15929675	0030234 : enzyme regulator activity			1			
serpina1a protein [*Mus Musculus*]	14602605	0030234 : enzyme regulator activity			1			
hypothetical protein [*Plasmodium Falciparum* 3D7]	23612616	-	-	-	3			
hypothetical protein [*Plasmodium Falciparum*]	23612848	-	-	-	4			
similar to macrophage galactose-type C-type lectin 2 [*Mus Musculus*]	38089618	-	-	-	12, 14			
hypothetical protein [*Plasmodium Yoelii Yoelii*]	23482237	-	-	-	13			
similar to retrovirus-related POL polyprotein [*Mus Musculus*]	38081301	-	-	-	36			
similar to spectrin alpha chain, brain (spectrin, non-erythrid alpha chain)	38074605	-	-	-	36			

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
similar to ribosomal protein L37 [*Rattus Norvegicus*]	27680290	-	-	-	36			
similar to hypothetical protein ORF-1137 [*Mus Musculus*]	38080223	-	-	-	36			
60S ribosomal protein L35a	3914537	-	-	-	42			
ribosomal protein L35a [*Rattus Norvegicus*]	10863991	0005198 : structural molecule activity; 0003676 : nucleic acid binding	0005622 : intracellular; 0005840 : ribosome; 0030529 : ribonucleoprotein complex	0006412 : protein biosynthesis	42			
similar to ribosomal protein L35a [*Mus Musculus*]	38084044	-	-	-	42			
unamed protein product [*Mus Musculus*]	26338025	0005198 : structural molecule activity; 0003676 : nucleic acid binding	0005622 : intracellular; 0005840 : ribosome	0006412 : protein biosynthesis	42			
similar to ribosomal protein L35a [*Mus Musculus*]	38075333	-	-	-	42			
ATP synthase subunit [*P. yoelii yoelii*	23482200	0005215 : transporter activity; 0016787 : hydrolase activity	0016469 : proton-transporting two-sector ATPase complex	0006091 : generation of precursor metabolites and energy; 0006793 : phosphorus metabolism; 0006811 : ion transport; 0006818 : hydrogen transport; 0009117 : nucleotide metabolism; 0051186 : cofactor metabolism	50			
merozoite surface protein 3 [*Plasmodium Falciparum*]	2281613	-	-	-	56			

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
protein kinase c inhibitor-like protein, putative [*Plasmodium Falciparum* 3D7]	23612841	-	-	-	62			
similar to 60S Ribsomal protein L29 (P23) [*Rattus Norvegicus*]	34866986	-	-	-	65			
unamed protein product [*Mus Musculus*]	26339834	-	-	-	66			
Pfj1 [*Plasmodium Falciparum*]	2351192	0005515 : protein binding; 0043167 : ion binding		0006457 : protein folding	66			
DNAJ protein [*P. falciparum* 3D7]	23510084	-	-	-	66			
hypothetical protein [*P. falciparum*]	23957743	-	-	-	73			
hypothetical protein [*P. yoelii yoelii*]	23482806	-	-	-	76			
hypothetical protein [*P. yoelii yoelii*]	23482975	-	-	-	79			
similar to 60S Ribsomal protein L29 (P23) [*Mus Musculus*]	38089196	-	-	-	80			
similar to 60S Ribsomal protein L29 (P23) [*Rattus Norvegicus*]	34881340	-	-	-	80			

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
unamed protein product [*Mus Musculus*]	26335691	0005515 : protein binding	0005737 : cytoplasm		81			
heparan sulfate 6-0-sulfotransferase 1 [*Mus Musculus*]	20845347	-	-	-	81			
similar to HRPAP20 short form [*Rattus Norvegicus*]	34867098	-	-	-	82			
similar to Rp17a protein [*Rattus Norvegicus*]	34868324	-	-	-	99			
hypothetical protein [*P. falciparum* 3D7]	23613213	-	-	-	99			
var -CD36 protein [P. falciparum]	3786199	0004871 : signal transducer activity	0016020 : membrane	0009405 : pathogenesis	99			
splicing factor, putative [*P. falciparum* 3D7]	23613195	-	-	-	99			
hemopexin	22022646	-	-	-				
Apoa4 protein [*Mus Musculus*]	6680702	0005215 : transporter activity; 0005515 : protein binding; 0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0050789 : regulation of biological process; 0006629 : lipid metabolism; 0006869 : lipid transport; 0006810 : transport; 0006066 : alcohol metabolism			19	
Ig heavy chain [*Mus Musculus*]	21927918	-	-	-			40	
baculoviral IAP repeat-containing 3; apoptosisi inhibitor 1[Mus]	6680696	0016874 : ligase activity; 0005515 : protein binding; 0043167 : ion binding	0005737 : cytoplasm; 0005622 : intracellular; 0005634 : nucleus	0008219 : cell death; 0050789 : regulation of biological process			35	

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
erythrocyte memebrane protein 1 [*P. falciparum*]	39842963	0004871 : signal transducer activity	0016020 : membrane	0009405 : pathogenesis			36	
apolipoprotein A-1 [*Mus Musculus*	2145141	0008289 : lipid binding	0005576 : extracellular region	0019538 : protein metabolism; 0006869 : lipid transport			42	
hypothetical protein [p yoelii]	23483254	-	-	-			47	
unamed protein product [*Mus Musculus*]	12862141	-	-	-			47	
similar to ERIC1 [*Rattus Norvegicus*]	34878530	-	-	-			48	
similar to pantothenate kinase 2; PANK2 [*Rattus Norvegicus*]	34858736	-	-	-			46	
putative yir3 protein [P yoelii]	23483075	-	-	-			46	
similar to hypothetical protein FLJ25333 [*Rattus Norvegicus*]	34852795	-	-	-			49	
brain stress early protein Gbi; rho GTPase activating protein 14; EST AI452337	27597098	-	-	-			50	
putative yir4 protein [P yoelii]	23482729	-	-	-			51	

Table 3.6 continued

Description	Accession	GO Slim Molecular Function	GO Slim Cellular Component	GO Slim Biological Process	Infected Gel Spot No.	Control Gel Spot No.	Infected (Depleted) Spots No.	Control (Depleted) Spot No.
erythrocyte memebrane protein 1 [*P. falciparum*]	40646494	0030246 : carbohydrate binding		0009405 : pathogenesis			54	
hypothetical protein [p yoelii yoelii]	23481917	-	-	-			52	

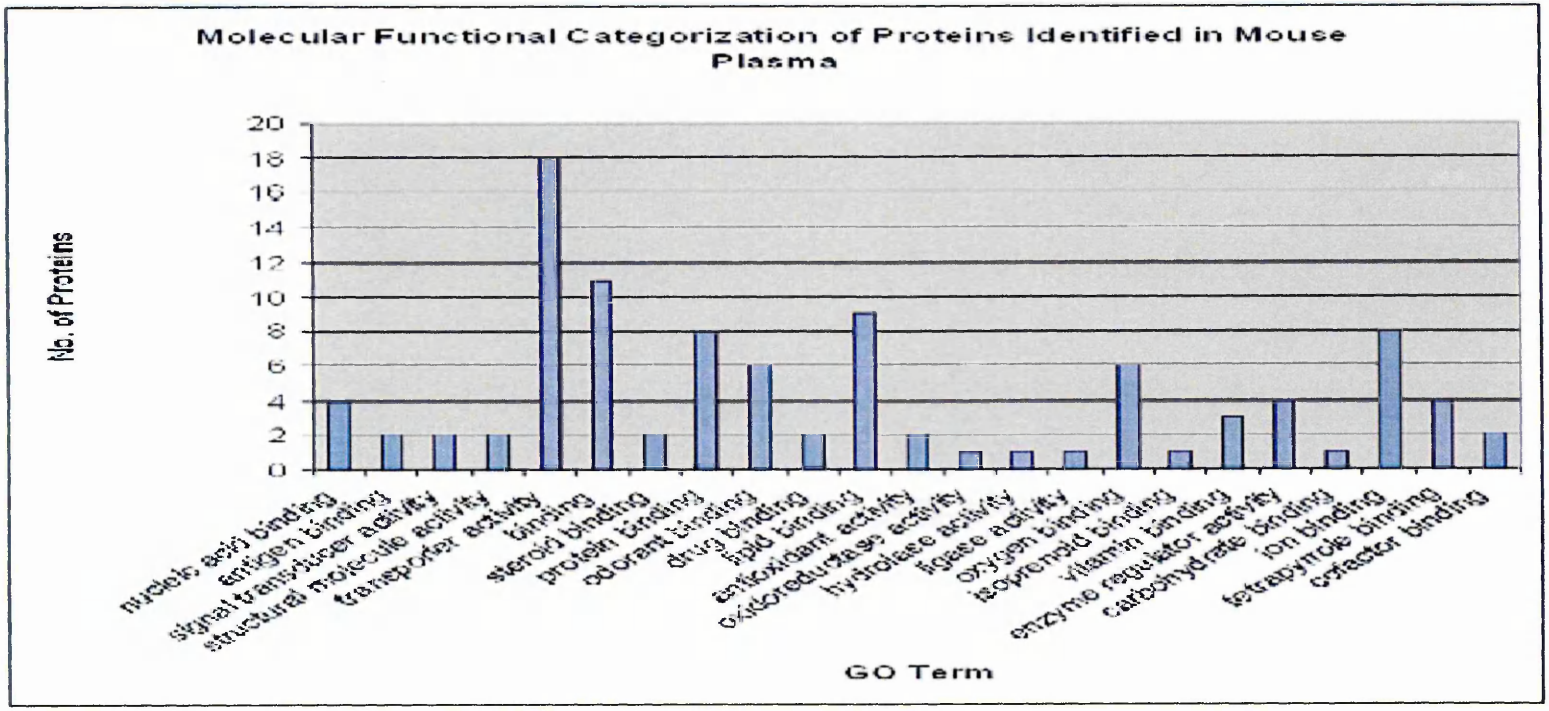

Figure 3.14 Functional Categorisation of 106 proteins identified based on gene ontology classification for Molecular Function The Gene Ontology (GO) numbers were derived from Protein Information Resource (PIR) -http://pir.georgetown.edu/ Batch retrieval tool. This tool converted the GI NCBInr accession numbers to GO numbers and then used GO slim to categorise the proteins. Another 47 identified proteins were unmatched to GO and not included in these graph.

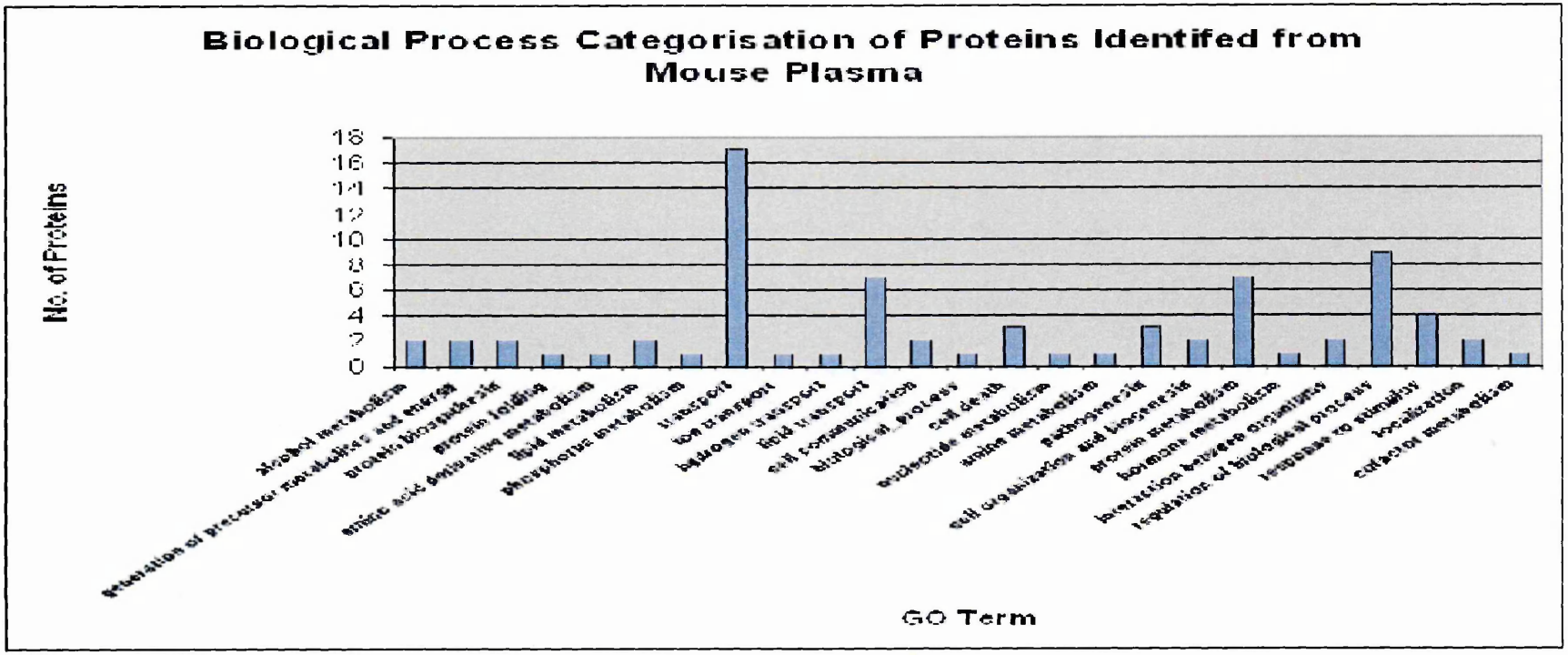

Figure 3.15 Functional Categorisation of 106 proteins identified based on gene ontology classification for Biological Process The Gene Ontology (GO) numbers were derived from Protein Information Resource (PIR) -http://pir.georgetown.edu/ Batch retrieval tool. This tool converted the GI NCBInr accession numbers to GO numbers and then used GO slim to categorise the proteins. Another 47 identified proteins were unmatched to GO and not included in these graph

3.4 Discussion

Animal models have been developed to provide insights into the pathogenesis of neurological complications of malaria. Mouse and rodent models have been studied most extensively to examine the function of the blood brain barrier in malaria. In this study the *Plasmodium berghei* ANKA mouse model includes neurological signs such as ataxia, hemiplegia and coma (Neill and Hunt, 1992), which are similar to the clinical features of human CM. However, in contrast to human CM, there is little evidence of sequestration of parasitised red blood cells (PRBCs) in the brain (Polder et al., 1992), but rather monocytes and non-PRBCSs predominate (Neill et al., 1993). Recent studies have found PRBCs in the cerebrum and cerebellum of infected mice (Hearn et al., 2000). Although the pathogenesis of the neurological complications of murine malaria appears to be more immunologically mediated (Polder et al., 1992, Hearn et al., 2000) than in human malaria, changes in protein components of mouse plasma may be indicative of disease related alterations.

Two-dimensional gel electrophoresis, though still the mainstay in proteomic research, has the main limitation of only having a dynamic range of 10^4. This creates difficulties when analysing crude plasma, which potentially has protein components ranging from the millimolar range to the low femtomolar range (Anderson and Anderson, 2002, Kenyon et al., 2002). Reducing the complexity of protein mixtures by depletion and fractionation of intact proteins greatly simplifies the task for 2-DE. There are essentially two patterns of depletion: depletion only of albumin or Ig or both, and depletion of the top-6 proteins, which are albumin, IgG, IgA, haptoglobin, alpha-1 anti-trypsin and transferrin. There is clear evidence that depletion makes it significantly more feasible to visualise, detect, and

then identify lower abundance proteins (Echan et al., 2005, Li et al., 2002a, Zolotarjova et al., 2005, Huang et al., 2005). To effectively identify low abundant proteins (LAPS) in plasma in this study, albumin, which together with another nine proteins accounts for almost 90% of the proteins in plasma, was removed. Although this made for better gel resolution, it was found that the many "new" spots detected after depletion were unmasked isoforms of high and medium-abundance proteins, rather than LAPS. Another counterbalancing problem noted by many authors is the non-target or inadvertent removal of other proteins, which could be due to peptides and proteins bound to the target proteins, especially albumin; cross-reactivity with the bound antibodies; or non-specific binding to the column or resin or dye (Granger et al., 2005). One solution for this would be pre-fractionating the sample prior to gel separation or to skip the gel separation and digest the fractions and then separate the resulting peptides by liquid chromatography. However, due to time constraints this was not explored in this study.

Protein alterations implicated in murine malaria include cell adhesion molecules such as P-selectin (Combes et al., 2004) and cytokines such as tumor necrosis factor (Grau and Lou, 1995). These proteins have been studied extensively using more conventional biochemical techniques, such as immunoassays (de Kossodo and Grau, 1993, Grau et al., 1989, Grau et al., 1990). Unfortunately, the dynamic range of the 2-DE approach, as well as the dynamic range of the stains used in this study is such that such proteins can theoretically only be detected after extensive work on developing a pre-fractionation strategy allowing enrichment of LAPs, which was not possible due to the limited time available. The dynamic range of the technique thus limited analysis to the more abundant acute-phase proteins, which are known markers of inflammatory disease including serum

albumin and various isoforms of apolipoproteins. Achieving the level of detection commensurate with immunoassay will only be possible on 2-DE using pre-fractionated plasma in which the major serum proteins (albumin and immunoglobulins) are substantially reduced, but not at the expense of the non-specific depletion of minor proteins. When this occurs, it will be possible to identify many subtle changes in protein expression, permitting protein enrichment to determine the mass spectrum and hence identity prior to development of specific immunoassays.

Proteins excised from preparative gels were processed for MALDI-ToF-MS. In the literature a number of cleavage enzymes have been used to digest proteins for MS. Here trypsin has been used as it displays a high level of specificity, cleaving the proteins explicitly after lysine or arginine residues. In addition trypsin is acknowledged to produce autoproteolytic fragments which can be used for internal calibration of the mass spectra (Beranova-Giorgianni and Desiderio, 2000). Trypsin digestion results in quite small peptides, which can be efficiently eluted from the gel matrix and which "fly" well in MALDI time of flight machines. All proteins were in-gel digested with trypsin, the resulting peptides were extracted from the gel piece and PMF was performed using MALDI-ToF-MS. PMF was selected as the preferred tool for protein identification in this study because the technique is fast and straight forward and therefore appropriate given the study duration and number of spots to be analysed.

Although MALDI is a very sensitive technique, not all spots investigated provided a statistically significant score on analysis. The Mowse Score reported by Mascot® is denoted by the expression -10*Log(P) where P is the probability the observed match is a random event. For *Rodents* proteins the scores were considered significant ($p<0.05$) if they

were greater than 63 while for *Alveoli* proteins scores greater than 57 were considered significant. The sequence coverage for proteins identified varied and results show high sequence coverage gave a high Mowse score.

Of the 157 proteins identified 54 were found to be uniquely expressed in the plasma from infected mice. One of these unique proteins was identified as similar to spectrin non-erythroid alpha chain brain protein (spot 36 infected gel map). This protein has recently been shown to correlate with severity of CM in Gabonese children (Guiyedi et al., 2007). Another of the proteins unique to plasma from infected mice was identified as similar to macrophage galcotose-type C-type lectin 2 (MGL2) (Spot 12 and 14). MGL2 is induced in diverse populations of activated macrophages, including peritoneal macrophages elicited during infection with the protozoan *Trypanosoma brucei brucei* or the Helminth *Taenia crassiceps* and alveolar macrophages elicited in a mouse model of allergic asthma (Raes et al., 2005). Raes *et al* also demonstrated that in vitro, interleukin-4 (IL-4) and IL-13 up-regulate MGL2 expression and that in vivo, induction of MGL2 is dependent on IL-4 receptor signaling. Moreover, expression of MGL on human monocytes is also up-regulated by IL-4. The study concluded that macrophage galactose-type C-type lectins represent novel surface markers for murine and human activated macrophages and could be important markers of immune response in murine malaria Another of the proteins identifies as unique to plasma drawn from infected mice is Baculoviral IAP repeat-containing protein 3 which is an apoptotic suppressor. This protein interacts with TNF receptor associated factors 1 and 2 to form an heteromeric complex, which is then recruited to the TNF receptor 2 (Liston et al., 1997). The presence of this protein could be as a

protective measure to stop apoptosis which has been shown to occur in murine liver cells during malarial infection (Guha et al., 2007, Guha et al., 2006).

There were 49 proteins uniquely expressed in the control gels. These included transthyretin (spot 72 control gel) which sometimes acts as a retinol carrier through an association with retinol binding protein (Sasaki et al., 1985). It is normally found at high concentrations in plasma and the fact that it was not detected in the plasma from infected mice would suggest that the levels in rodent malaria are too low to detect or some protein modification had occurred. Transthyretin is an acute phase protein that decreases during an acute phase response (Schreiber et al., 1989). Together with other acute phase proteins such as transferrin, and albumin, these proteins that decrease during an acute phase response have no apparent immune function (Schreiber et al., 1989). Their main role is to transport nutrients and, therefore, their reduction during infection and inflammation may lower the concentration of specific nutrients. The serum concentration of retinol, the alcohol form of vitamin A, decreases during malarial infections. This reduction has been characterised as a direct consequence of the inflammatory response to *Plasmodium* infections (Thurnham and Singkamani, 1991, Tabone et al., 1992). Previously, several proteins have been associated with a reduction in retinol (Filteau et al., 1993, Das et al., 1996, Friis et al., 1997). The proteomic results suggest that transthyretin may be a useful predictor of plasma retinol during malarial infection. This information would be useful because the inclusion of a measure of the acute phase response would help interpret plasma retinol concentrations during malarial infection (Thurnham and Singkamani, 1991, Filteau et al., 1993).

Proteins found to be common to infected and non-infected plasma were mainly variants of albumin and apolipoproteins and were identified to many spots in different regions of the

2-DE gels suggesting that these proteins had gone through some form of post translational modification.

3.5 Conclusion

This study helped to validate the technique of proteomics for characterising differentially expressed LAPS in different stages of the severe forms of murine malaria. It also highlighted several problems in applying this technology to plasma due to the limitations of 2-DE that is known to have a systemic bias against very large, very small and LAPs (Wilkins et al., 2006). This only leaves proteins expressed at medium or high levels from the 2-DE. Using 2-DE together with pre-fractionation techniques could greatly reduce the problems associated with this technique (Tam et al., 2004) increasing the amount of information that can be derived from this technique. Some of these problems are addressed in the next chapter which deals with characterisation of LAPs in plasma samples from patients.

4 CSF and Plasma Proteomes from Archived Samples from CM, ABM and EN

4.1 Introduction

The successful completion of the human and *Plasmodium falciparum* genome projects has created an opportunity to increase our understanding of the molecular basis of severe malaria. However, a comprehensive understanding of the dynamic protein pathways involved in a) normal and disease states and b) response to medical treatment, is required to improve the outcome.

Traditional analysis of biomarkers has concentrated on identifying one marker for a particular disease. However, there is now a general agreement, supported by a statistical argument, that for complex disease states a panel of independent but disease-related proteins, considered in an aggregate, would be more relevant as they should be less prone to the influence of genetic and environmental 'noise' than is the case for a single marker protein (Anderson, 2005). Importantly proteomics has the power to identify such panels of proteins. The 2-DE proteomic approach has been most successfully used to identify protein profiles in various disease states. For example, it was demonstrated that leukaemias could be classified into their different subtypes using 2-DE (Hanash et al., 2002). This methodology has also been applied in the investigation of infectious diseases, diseases of the central nervous system, heart and cancer (reviewed in (Zolg and Langen, 2004, Engidawork and Lubec, 2001, Rohlff, 2000, Banks et al., 2000, Tyers and Mann, 2003, Petricoin et al., 2002e, Hanash, 2003, Bichsel et al., 2001, Alaiya et al., 2000, Fountoulakis and Lahm, 1998, Fountoulakis, 2001). This chapter describes the application of methods developed in chapter 3 to create gel maps of plasma and CSF collected from children diagnosed with cerebral malaria (CM) and compared these findings to patients with undiagnosed encephalopathy (EN) and acute bacterial meningitis (ABM), in an effort to try and understand differences in protein expression. The diseases

pheonotypes were chosen based on results of previous analyses of CSF samples using NMR (Figure 1.3) spectrum of CSF from patients with CM is considerably different from children with meningitis and other neurological conditions e.g. stroke and multiple sclerosis. Table 4.1 highlights some of the differences between the 3 groups. We have employed the 2-DE approach to try and define multiple markers for severe malarial disease.

4.2 Materials and Methods

4.2.1 Patient Samples

Archived plasma and CSF samples(stored at -80°C in heparinised tubes) from children diagnosed with CM, EN and ABM were collected to study the changes in CSF using techniques other than those described in this study (SSC Protocol No. 480 "Mechanisms of neurological damage in Kenyan children with cerebral malaria and acute bacterial meningitis" Newton CRJC et al). The analytical methods described in this thesis are more sensitive and selective and were used in place of those described earlier in the SSC Protocol No. 480. The change in analytical methods used was described in SCC protocol No. 772 which was approved during the 103rd KEMRI/National Ethical Review Committee meeting held on 3rd June 2003. Based on results of previous analyses of CSF samples using NMR (**Fig. 1.1**) spectrum of CSF from patients with CM is considerably different from children with meningitis and other neurological conditions e.g. stroke and multiple sclerosis, suggesting presence of compounds uniquely expresses in CM. Due to the invasive nature of sample collection it would not be ethical to have samples collected from healthy children, therefore samples from children with a diagnosis of ABM and EN were used as control groups. **Table 4.1** gives a clinical comparison of these disease phenotypes and **table 4.2** gives demographic data of patients used in this study.

Table 4.1 Clinical comparisons of Cerebral Malaria (CM), Acute Bacterial Meningitis (ABM) and Undiagnosed Encephalopathy (EN)

	CM	*ABM*	*EN*
Aetiology	Parasite in RBC	Bacteria	Viruses
Pathology	Infected RBC confined to vessels	Inflammation of meninges with vascilitis	Infection of parenchyma of the brain
Breakdown of BBB	Moderate	Severe	Mild
Clinical features	Coma	Variable conscious level from fully conscious to deep coma	Variable conscious level from fully conscious to deep coma
Seizures	>80% (Idro et al., 2006)	53.8% (Waruiru et al., 1996)	19.7% (Waruiru et al., 1996)
Outcome Sequelae	11%	10-30%	20%

Table 4.2 Clinical characteristics of study patients

	Cerebral Malaria (CM)	Acute Bacterial Meningitis (ABM)	Undiagnosed Encephalopathy (EN)
Age[3]	*34 (17.5)*	*26 (23.3)*	*37 (28.6)*
Sex (m/f)	*6/6*	*4/8*	*4/8*
Duration of illness before admission to hospital[4]	*2.75 (1.71)*	*10.83 (27.94)*	*2.45 (1.81)*
Number with Seizure	*8*	*11*	*8*
Number with Coma	*9*	*5*	*6*
Hb (SD)	*8.42(2.47)*	*9 (1.89)*	*8.97 (2.54)*
Duration in Hospital[5]	*4.58 (5.9)*	*4.67 (3.06)*	*20.5 (43.1)*
Moratlity	*25%*	*17%*	*33%*

[3] Average age in months (SD)

[4] Average duration in days (SD)

[5] Average duration in days (SD)

4.2.2 S ample Preparation

To minimise protein loss, plasma samples were used with no prior sample preparation. However, CSF contains a high salt concentration (>150 mmol/L) and a low protein concentration (the CSF/serum ratio of protein concentration is 4×10^{-3}, 200–700 μg protein/mL).(Yuan and Desiderio, 2005a). This causes interference with electrophoretic separation of proteins because of the high electrical current that is carried by the salt load. This in turn reduces the efficiency of the 2-DE. Therefore, CSF samples were prepared as follows: the CSF samples were desalted using Microcon YM-3 centrifugal filter units (Millipore, USA) according to manufacturer's instructions using water as the buffer of choice. These units filter out anything with less than 3kDa and also enabled the concentration of the CSF samples.

4.2.3 P rotein Determination Using the Bradford Assay

Protein concentration of diluted plasma (1:100 v/v with water) and desalted CSF was determined as described in chapter 2 section 2.1.2.1

4.2.4 Tw o Dimensional Gel Electrophoresis

Following determination of the protein concentration (Section 4.2.3), proteins were separated using 2-DE as describe in chapter 2 sections 2.1.8.2-2.1.8.4. Analytical gels used to provide master maps were loaded with 75μg of protein and preparative gels to be stained with Coomassie stain were loaded with 300μg of protein. Due to the low amount of protein content in CSF, preparative gels were not prepared and analytical gels stained with silver stain were used for MS analysis.

4.2.5 P rotein visualisation and image analysis

Analytical gels were silver stained as described in chapter 2 section 2.1.8.5 and preparative gels stained with Coomassie as described in chapter 2 section 2.1.8.5 The stained gels were scanned using GS-710 Imaging Densitometer (BioRad) and analysed using PDQuest®

software version 7.0 (Biorad) (chapter 2 section 2.1.2.7) and Progenesis® 200 software (Nonlinear Dynamics, Newcastle upon Tyne, UK) (chapter 2 section 2.1.8.5). The software enabled the matching of identical spots in serial gels and normalisation of gels to compensate for non-expression related variations in protein spot intensity.

4.2.6 Try ptic in-gel digestion

For plasma, spots cut from gels stained with Coomassie were digested as described in chapter section 2.2.3.1. For CSF, spots cut from silver stained gels were digested using a modified method described by (Terry et al., 2004). The silver stained spots were washed twice in 50% ACN in 100mM ABC for five minutes. The spots were then dehydrated using 100% ACN until they turned opaque white, followed by evaporation of excess ACN in a vacuum concentrator. Digestion buffer (5-10µl of 10µg/ml modified sequencing grade trypsin (Sigma-Aldrich) in 25mM ABC), was added to the dried gel pieces and incubated overnight at 37°C. Resulting peptides were extracted by first incubating the digestin 7µl of distilled water and then adding 30%ACN in 0.1% TFA and sonicating for 30 minutes. The digest was then vortex mixed and centrifuged for 2 minutes. The supernatant was transferred and vacuum concentrated to about 2µl.

4.2.7 M ass Spectrometry

4.2.7.1 Characterisation of protein spots by MALDI-ToF-MS

All spots digests were analysed using MALDI-ToF as described in chapter 2 section 2.3.9.1.

4.2.7.2 Characterisation of Protein Spots by RP-LC-MS/MS

To increase the number of proteins identified, the remaining spot digest was diluted using 15µl of a 50% acetonitrile/5% formic acid v/v mixture and analysed by tandem mass spectrometry using a LCQ Deca XpPlus mass spectrometer (ThermoFinnigan, USA) equipped with an electrospray source connected to a reversed phase Thermo Hypersil-Keystone biobasic C18

column (0.18 x 100mm) as follows: after sample injection, peptides were eluted on a 60 min gradient 2% to 50% ACN at a flow rate of 500ul/min. To maintain a stable spray, the spray voltage was adjusted to (1.5-1.8kV). The capillary temperature was set at 250°C and the normalised collision energies were set at 35% for MS/MS. The mass spectrometer was operated on a data-dependent "triple play" mode where the 3 most intense ions in the full scan were subjected to a zoom scan followed by MS/MS. Dynamic exclusion set at 0.5 min was used to obtain MS/MS spectra from any co-eluting peptides.

4.2.8 Data base Searching and Protein Categorisation

Mass lists generated from MALDI-ToF were searched against the NCBInr database using the search parameters described in chapter 3 section 3.2.7. MS/MS data generated from the LC-MS/MS experiment were searched using TurboSequest® described in detail in chapter 2 section 2.3.5.4. The tandem mass spectra were searched against the NCBInr database as well as the *Plasmodium falciparum* database downloaded from the Sanger Institute. Proteins were accepted if they were identified with 2 or more peptides and only if they were identified from both the subset database as well as from searching the entire NCBInr database. Protein cataloguing was done according to the GO terms (chapter 2 section 2.3.8.1) for molecular function, cellular component and biological process using the tool provided by PIR as described fully in chapter 2 section 2.3.9.5

4.3 Results

4.3.1 P rotein Determination Using the Bradford Assay

The mean plasma protein concentrations were 43.44 g/L (s.d. 19.51), 51.17 g/L (s.d. 19.72) and 47.44 g/L (s.d. 15.83) for CM, ABM, and EN respectively. Table 4.3 gives CSF total protein contents and compares them to those found in literature.

Table 4.3 Total protein content for CSF samples

Cerebral Malaria	*Acute Bacterial Meningitis*	*Encephalitis*	*Source*
0.53 g/L *(0.54)* *n=12*	*3.34g/L* *(3.02)* *n=12*	*0.23 g/L* *(0.09)* *n=12*	*My results*[6]
0.4g/L	-	*0.5g/L*	*(Jakka et al., 2006)*
0.19g/L	*1.56g/L*	-	*(Berkley et al., 1999)*

[6] Results are given as mean (SD)

4.3.2 P rotein Separation

In order to compare protein expression profiles of plasma and CSF, representative samples collected from children diagnosed with CM, ABM and EN were applied to 2-DE and the proteins visualised by silver stain and coomassie stain. To prepare gel maps of proteins PDQuest® 2D software was used. Gel maps for plasma were created from both silver and Coomassie stained gels. Figure 4.1 shows master gel maps for plasma from CM, ABM and EN respectively from the coomassie stained gels. The master gel maps for CSF are shown on Figures 4.2. The master gels were prepared by analysing duplicate gels of 12 patients and spots matched in 75% of the gels included into the master gel. Figures 4.3-4.4 represents a composite gel showing differences identified between plasma gel maps of CM compared to ABM and CM compared to EN respectively Figures 4.5-4.6 shows a composite gel showing differences between CSF gel maps of CM compared to ABM and CM compared to EN.

Semi-quantitative analysis of the gels was performed using the Progenesis® 200 software which highlighted any spots that may have been up or down regulated. Figures 4.7 and 4.8 show the spots that had differential expression on the CM plasma gels and CM CSF gels respectively.

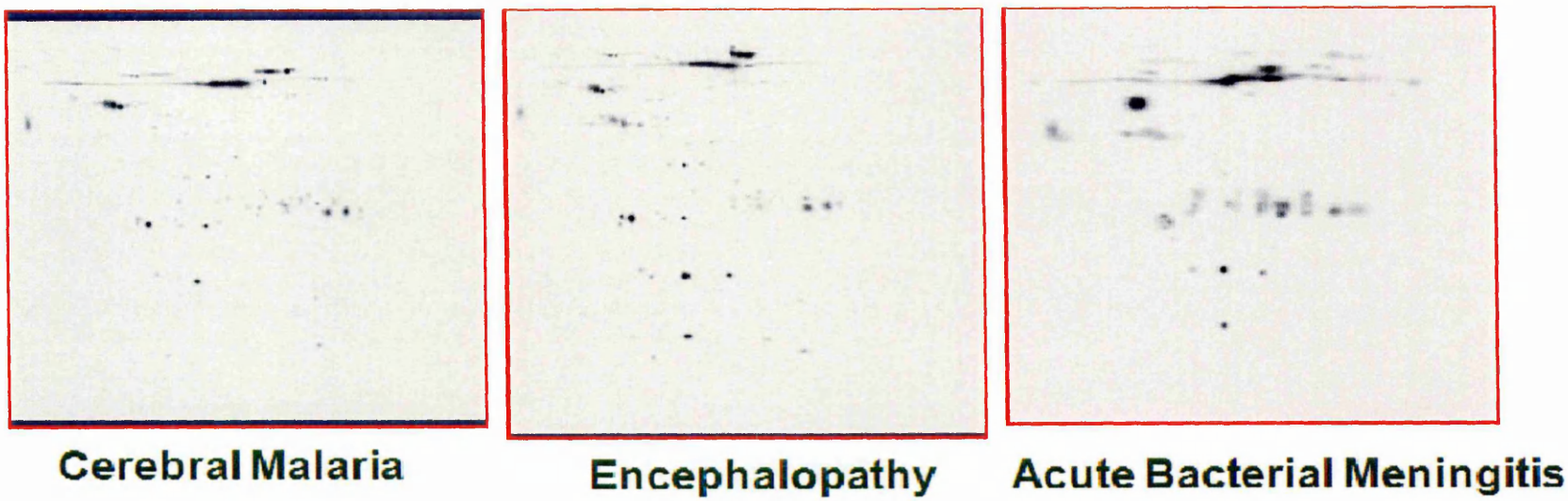

Figure 4.1 Master gels maps created using PDQuest® for plasma collected from children diagnosed with CM, EN and ABM. The gel maps are synthetic images generated using replicate gels of a match-set. The replicates (n=24) were from duplicate gels from 12 patients per disease group and spots were included if found in 75% of the gels. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SD-PAGE (12.5% T constant).

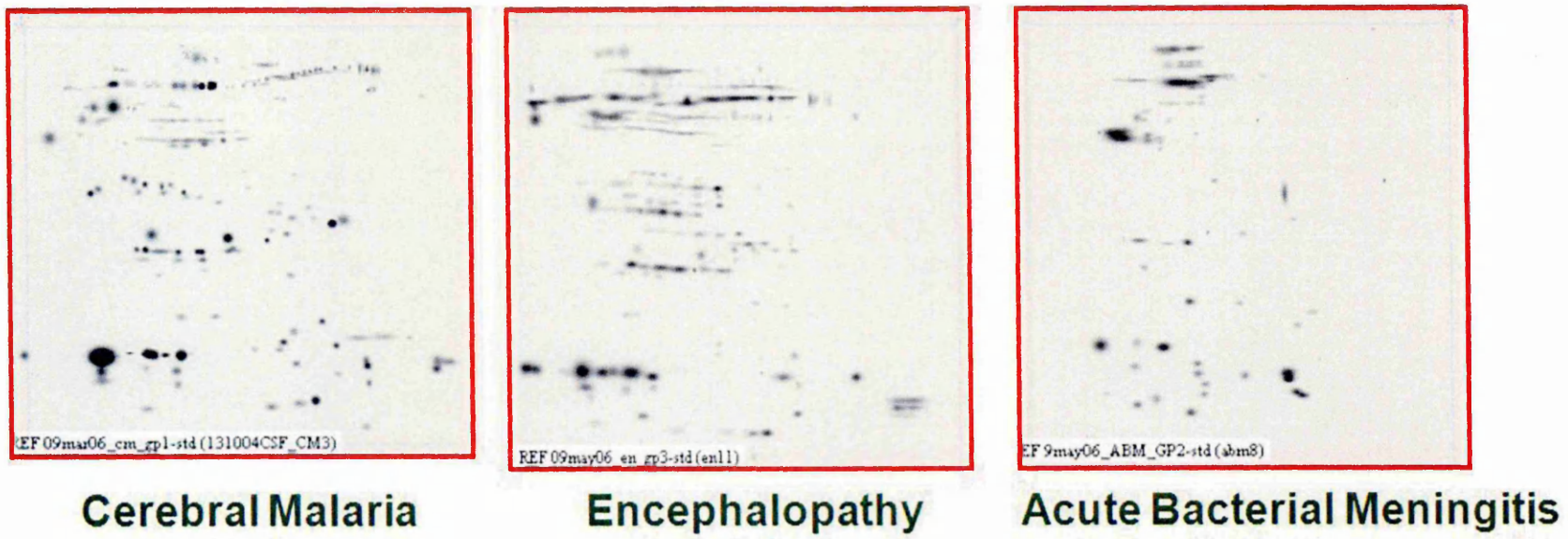

Figure 4.2 Master gels maps created using PDQuest® for CSF collected from children diagnosed with CM, EN and ABM. The gel maps are synthetic images generated using replicate gels of a match-set. The replicates (n=24) were from duplicate gels from 12 patients per disease group and spots were included if found in 75% of the gels. The first dimension was carried out using commercial pre-cut IPG 3-10 strips 130mm (24 h at 3000Vmax). Second dimension was run using vertical SDS-PAGE (12.5% T constant).

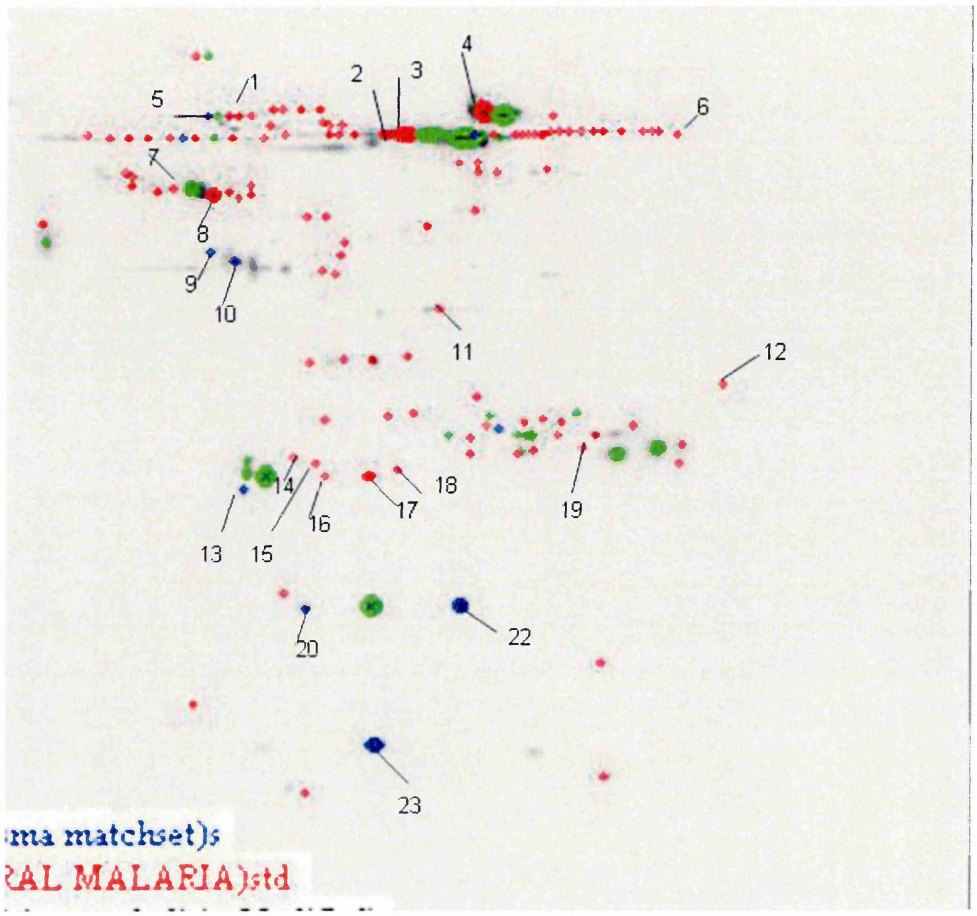

Figure 4.3 Composite gel showing the differences between gels of plasma collected from children diagnosed with CM and ABM. Composite gels were created using the comparison tool in PDQuest®. The composite gel demonstrates qualitative differences in the CM gel map compared to the ABM gel map. Red spots are spots found in the CM gel map and blue spots are those from the ABM gel map. Green spots depict spots found in both. Some of the numbered spots were identified using MS and are listed on **Table 4.4**

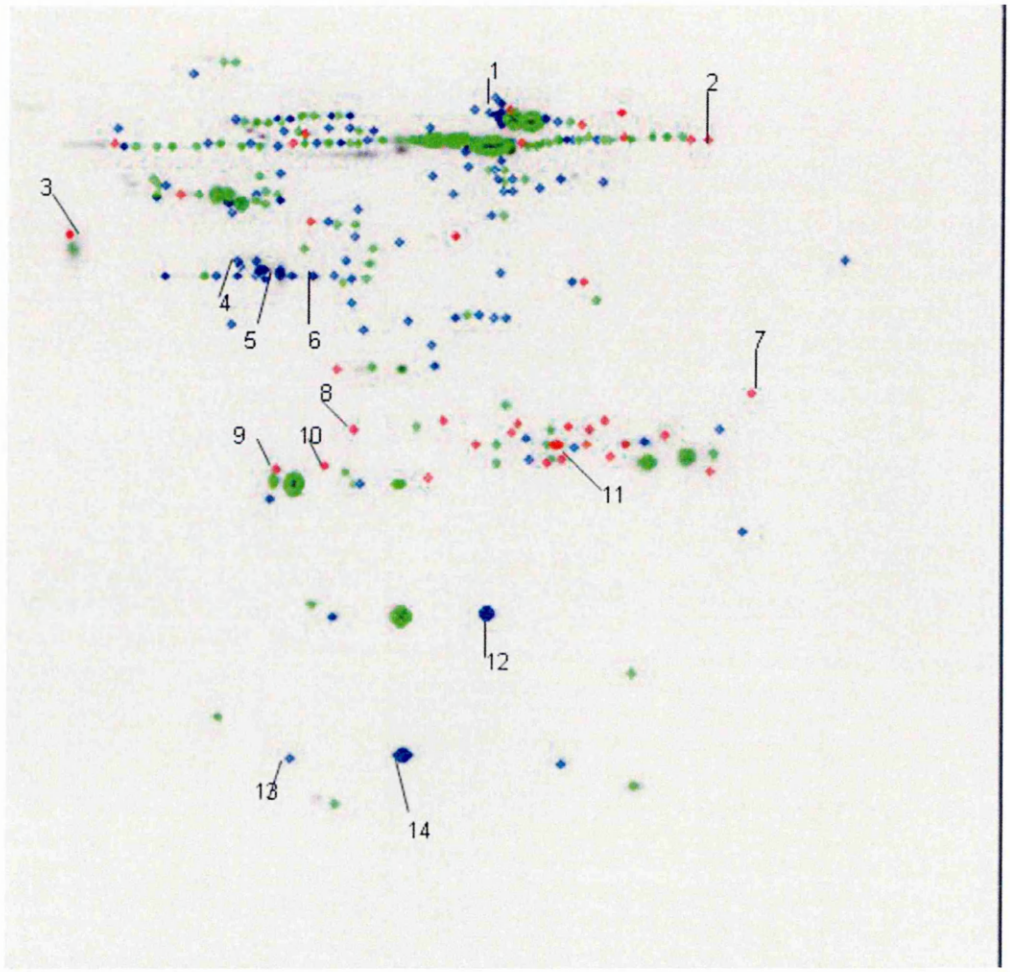

Figure 4.4 Composite gel showing the differences between gels of plasma collected from children diagnosed with CM and EN. Composite gels were created using the comparison tool in PDQuest®. The composite gel demonstrates qualitative differences in the CM gel map compared to the EN gel map. Red spots are spots found in the CM gel map and blue spots are those from the EN gel map. Green spots depict spots found in both. Spots successfully identified using MS are listed on **Table 4.5**

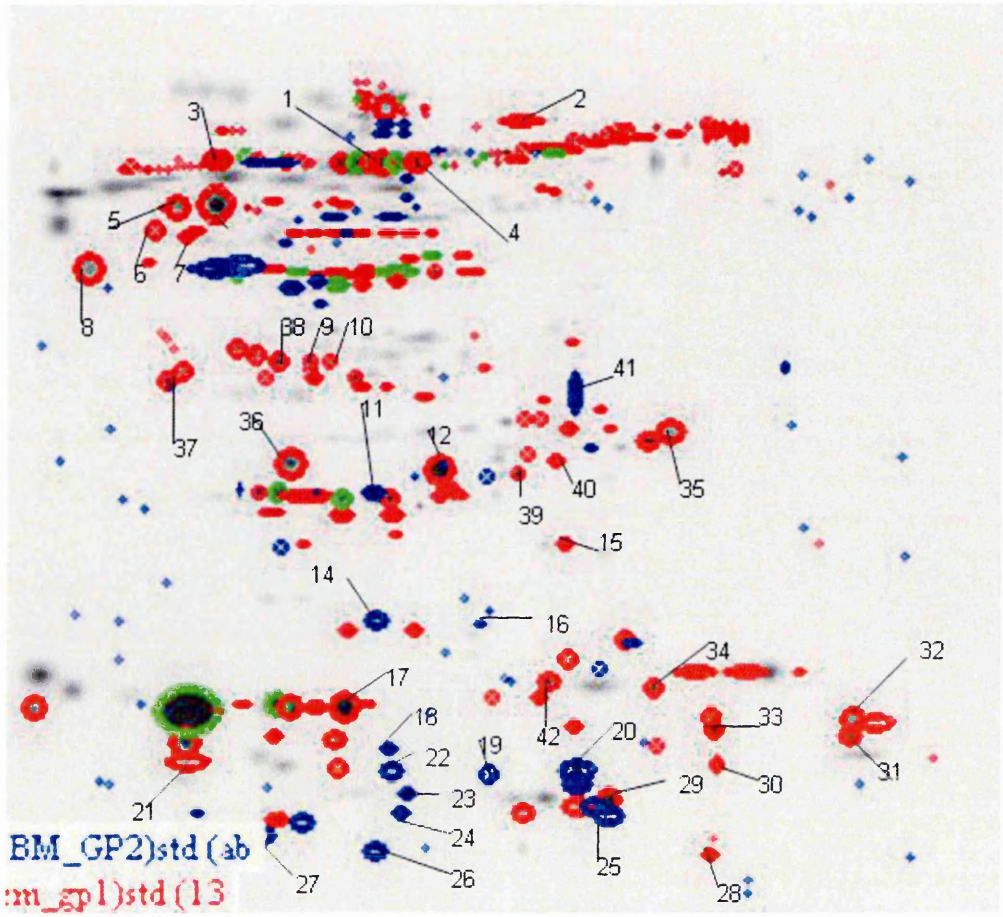

Figure 4.5 Composite gel showing the differences between gels of CSF collected from children diagnosed with CM and ABM. Composite gels were created using the comparison tool in PDQuest®. The composite gel demonstrates qualitative differences in the CM gel map compared to the ABM gel map. Red spots are spots found in the CM gel map and blue spots are those from the ABM gel map. Green spots depict spots found in both. Spots successfully identified using MS are listed on **Table 4.6**

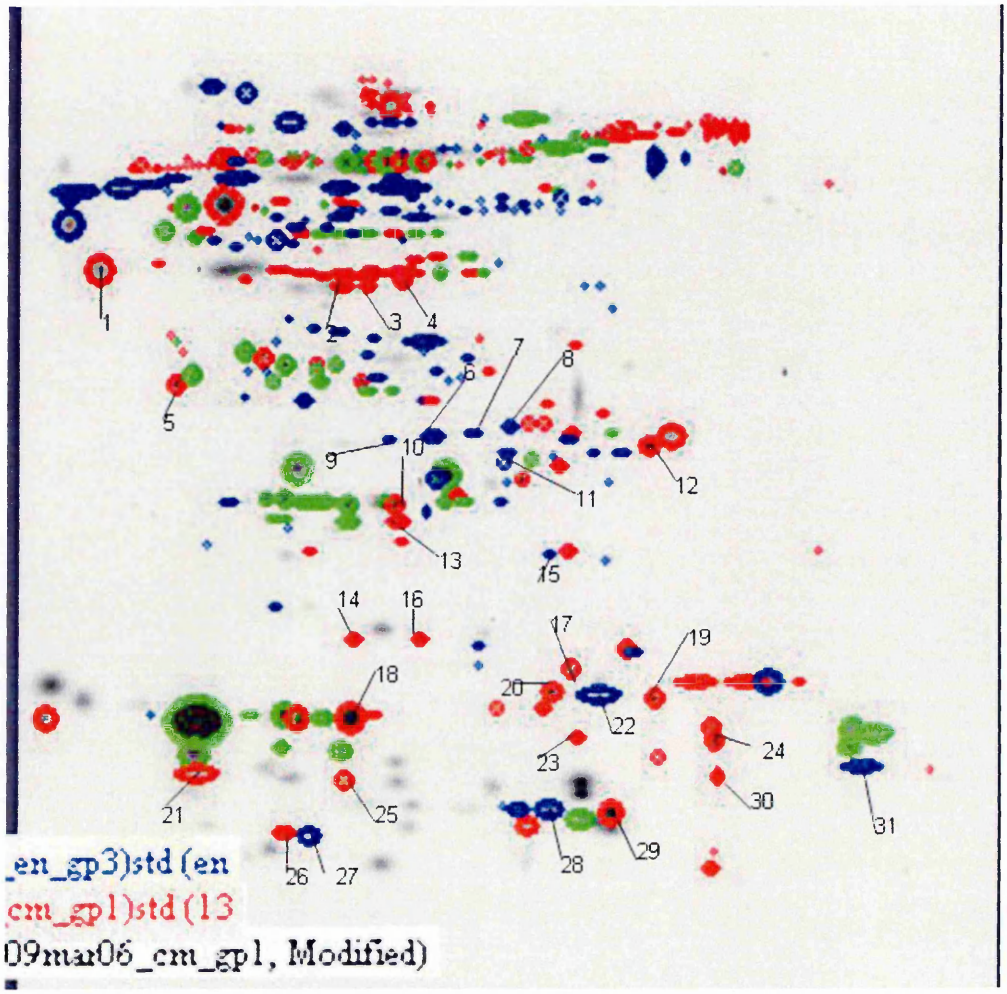

Figure 4.6 Composite gel showing the differences between gels of CSF collected from children diagnosed with CM and EN. Composite gels were created using the comparison tool in PDQuest®. The composite gel demonstrates qualitative differences in the CM gel map compared to the EN gel map. Red spots are spots found in the CM gel map and blue spots are those from the EN gel map. Green spots depict spots found in both. Spots successfully identified using MS are listed on **Table 4.7**

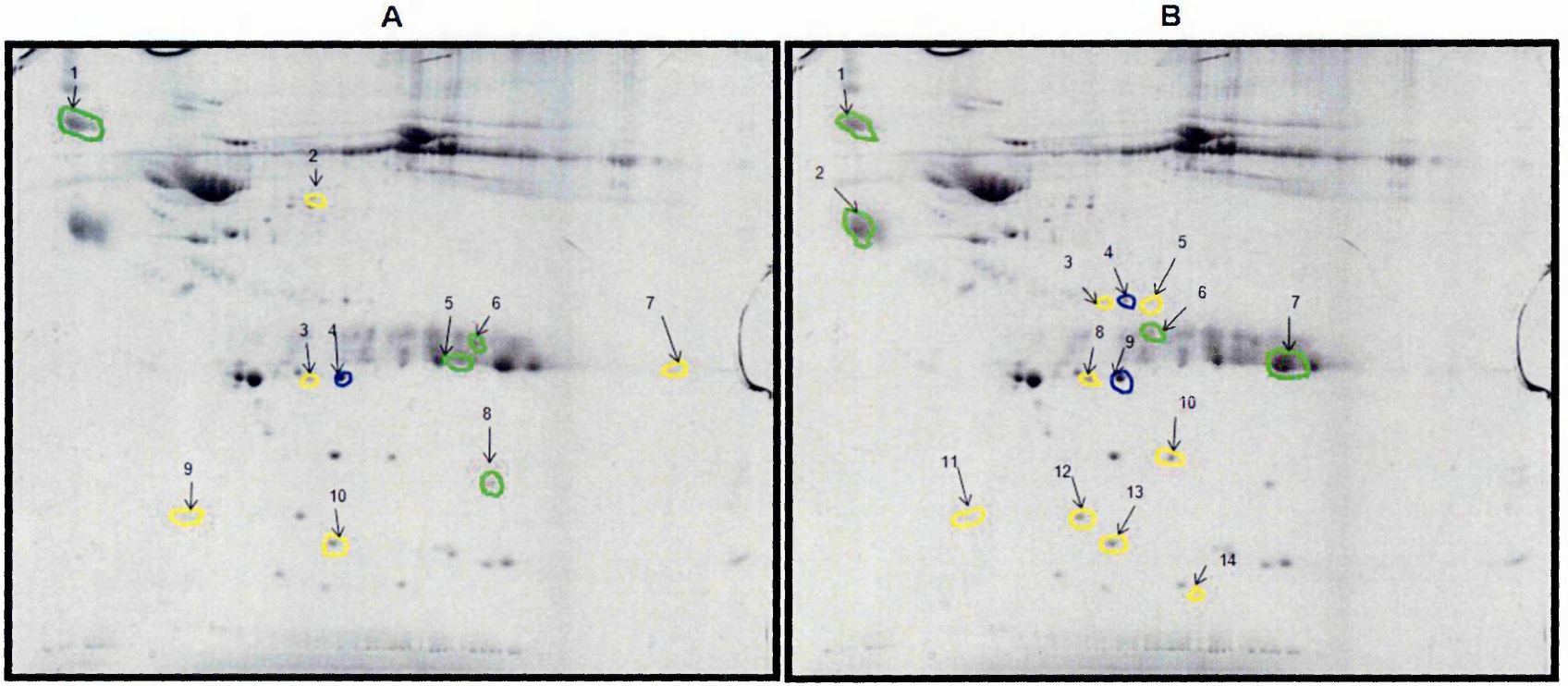

Figure 4.7 Representative gel of CM plasma showing differential expression of spots. (A) shows differences compared to ABM and (B) shows differences compared to EN. The green spots show normalised spot volumes with an up-regulation of a factor of 2 or more and the yellow spots show down regulation of normalised spot volume by a factor of 2 or more. The blue spots are unique to the gel.

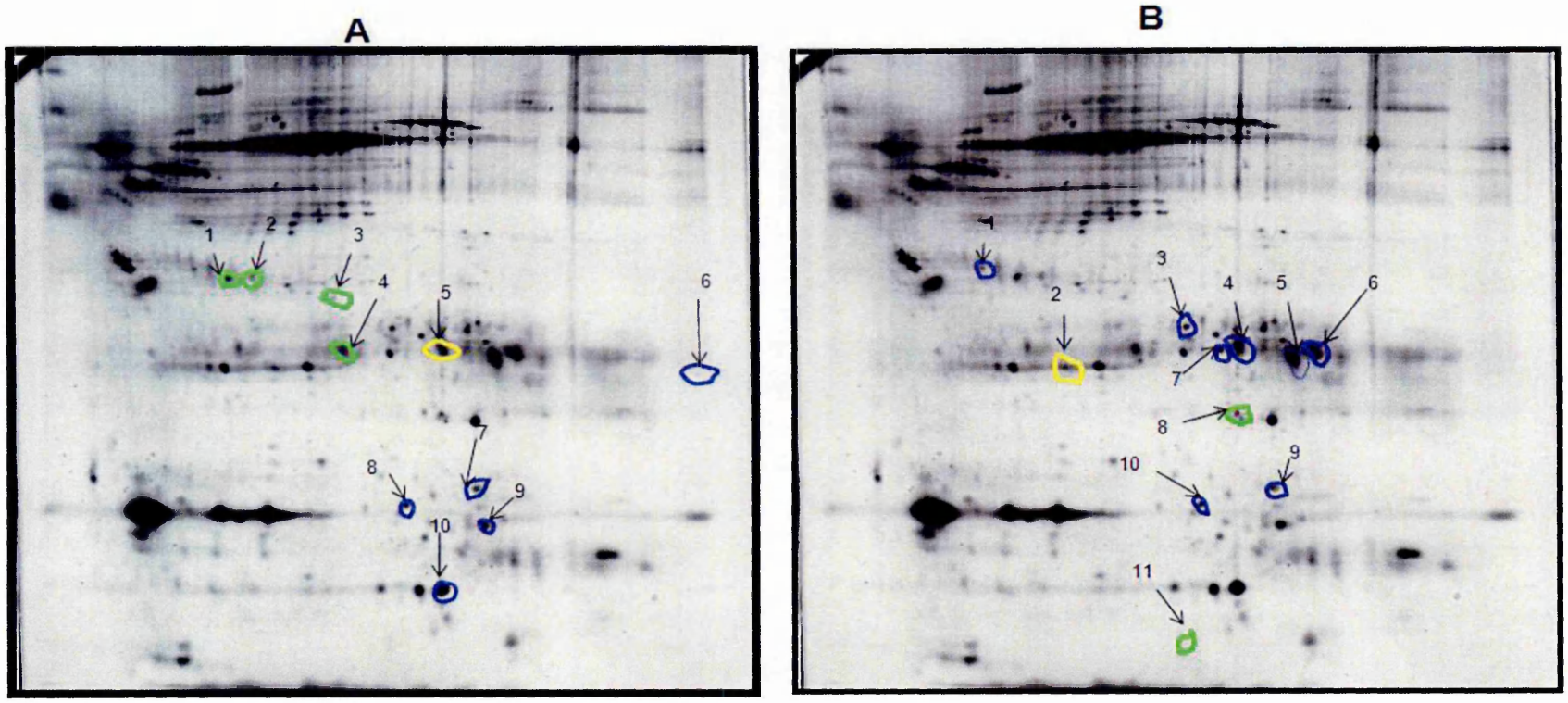

Figure 4.8 Representative gel of CM CSF showing differential expression of spots. (A) shows differences compared to ABM and (B) shows differences compared to EN. The green spots show normalised spot volumes with an up-regulation of a factor of 2 or more and the yellow spots show down regulation of normalised spot volume by a factor of 2 or more. The blue spots are unique to the gel

4.3.3 P rotein Identification by MS

Preparative gels for plasma were stained with Coomassie however due to the low protein content CSF protein spots were cut out from the analytical silver stained gels. **Tables 4.5-4.8** list protein identifications of the spots of interest labelled on **Figures 4.3-4.6** and **Tables 4.9 and 4.10** list proteins identified from spots labelled on **Figures 4.7 and 4.8.** Of the Coomassie stained gels 22% -42% of the spots on each gel were identified, while of the silver stained gels approximately, 10% of the spots per gel were identified. Irrespective of the staining method protein identification from ABM gels were very low (22% for Coomassie and 9% for silver stain).

4.3.4 Characteri sation and Cataloguing of Proteins

Protein functions were identified using the PIR batch retrieval tool and catalogued according to their gene ontology (GO) number. The gene ontologies gave the molecular function, the cellular component and biological processes of the proteins that had GO numbers and are part of **Tables 4.4-4.9.**

Table 4.4 List of proteins identified using MS from spots of interest shown on Figure 4.3 (composite plasma gel of CM compared to ABM) Proteins where identified using both MALDI-ToF and ESI-MS/MS and sequence coverage given in brackets is from ESI-MS/MS. Red indicates spot was found in CM only, blue in ABM only and green in both.

Spot No	Description	Accession	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage (%)
4	4557871	Transferrin	0043167 : iron binding	0005576 : extracellular region; 0016020 : membrane; 0045177 : apical part of cell; 0016023 : cytoplasmic membrane-bound vesicle	0019725 : cell homeostasis; 0006811 : ion transport	22(9.89)
7	15990507	Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor)	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	25
	110350939	Alpha-1 antitrypsin variant	0030234 : enzyme regulator activity			25
	177836	Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor)	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	25
	1703025	A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	- (16.75)
	4507871	vav 2 oncogene; Oncogene VAV2 [Homo sapiens]	no match			- (4.21)
8	28193184	Full-length cDNA clone CS0DM003YH10 of Fetal liver of Homo sapiens	0030234 : enzyme regulator activity			33
	1703025	A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	- (24.4)
17	3126979	Sorting nexin-3 (Protein SDP3)	0008289 : lipid binding; 0005515 : protein binding	0005737 : cytoplasm	0015031 : protein transport; 0007154 : cell communication; 0006810 : transport; 0016044 : membrane organization and biogenesis; 0016192 : vesicle-mediated transport	32
23	3608469	immunoglobulin heavy chain variable region				71

Table 4.5 List of proteins identified using MS from spots of interest shown on Figure 4.4 (composite plasma gel of CM compared to EN) Proteins where identified using both MALDI-ToF and ESI-MS/MS and sequence coverage given in brackets is from ESI-MS/MS. Red indicates spot was found in CM only and blue in EN only

Spot No	Description	Accession	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage (%)
1	C Chain C, Human Serum Transferrin, Recombinant N-Terminal Lobe, Apo Form	4389232	0043167 : ion binding	0005576 : extracellular region; 0005768 : endosome; 0016023 : cytoplasmic membrane-bound vesicle	0006810 : transport; 0019725 : cell homeostasis; 0006811 : ion transport	- (7.42)
	transferrin [Homo sapiens]	4557871	0043167 : ion binding	0005576 : extracellular region; 0016020 : membrane; 0045177 : apical part of cell; 0016023 : cytoplasmic membrane-bound vesicle	0019725 : cell homeostasis; 0006811 : ion transport	- (3.15)
3	orosomucoid 2; alpha-1-acid glycoprotein, type 2 [Homo sapiens]	4505529	0005488 : binding	0005576 : extracellular region	0050896 : response to stimulus	11.94
	orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1); a	9257232	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	15.42
5	interleukin 7 receptor, isoform CRA_b	119576329	-	-	-	60
6	fibrinogen gamma	223170	0043167 : ion binding; 0005515 : protein binding	0005576 : extracellular region; 0043234 : protein complex	0050817 : coagulation; 0050896 : response to stimulus; 0007165 : signal transduction; 0009987 : cellular process; 0019538 : protein metabolism; 0008150 : biological_process; 0008283 : cell proliferation; 0050789 : regulation of biological process	32
	fibrinogen gamma chain, isoform CRA_a	119625310	-	-	-	34
	F Chain F, Crystal Structure Of Crosslinked Fragment D	2781218	no match	no match	no match	- (11.29)
11	M-phase phosphoprotein 1; mitotic kinesin-like protein [Homo sapiens]	7705348	no match	no match	no match	- (2.42)
14	albumin, isoform CRA_g	119626070	-	-	-	35

Table 4.6 List of proteins identified using MS from spots of interest shown on Figure 4.5 (composite CSF gel of CM compared to ABM) Proteins where identified ESI-MS/MS. Red indicates spot was found in CM only.

Spot No	Description	Accession	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage (%)
9	transmembrane tryptase preproprotein; mast cell tryptase; mast cell pro	6912728	no match	no match	no match	4.79
	ras homolog gene family, member B; RhoB; Aplysia RAS-related homolog 6	4757764	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0005515 : protein binding	0016020 : membrane; 0005622 : intracellular; 0005768 : endosome; 0005634 : nucleus	0015031 : protein transport; 0007275 : development; 0008219 : cell death; 0007165 : signal transduction; 0007155 : cell adhesion; 0006996 : organelle organization and biogenesis; 0016192 : vesicle-mediated transport; 0046907 : intracellular transport; 0006810 : transport; 0007049 : cell cycle; 0050789 : regulation of biological process	7.14
10	similar to KIAA0215 gene product; similar to	27669054	no match	no match	no match	2.93
	mitochondrial ribosomal protein L32 [Homo sapiens]	13994261	0045182 : translation regulator activity; 0005198 : structural molecule activity	0005739 : mitochondrion; 0016020 : membrane; 0030529 : ribonucleoprotein complex; 0005840 : ribosome	0006412 : protein biosynthesis; 0015031 : protein transport; 0046907 : intracellular transport	4.66
	transmembrane tryptase preproprotein; mast cell tryptase; mast cell pro	6912728	no match	no match	no match	4.79
	T43453 hypothetical protein DKFZp434E0418.1 - human (fragmen	11359965	no match	no match	no match	5.26
	isocitrate dehydrogenase 3 (NAD+) alpha [Homo sapiens]	5031777	0016491 : oxidoreductase activity	0005739 : mitochondrion	0008152 : metabolism; 0005975 : carbohydrate metabolism; 0006091 : generation of precursor metabolites and energy; 0051186 : cofactor metabolism	3.55
	A8B3_HUMAN Potential phospholipid-transporting ATPase IK	8134319	no match	no match	no match	1.38
	PF14_0677 PF14_0677 RNA 3'-Terminal Phosphate Cyclase-like protein, putative 22874540:22875943 forw	PF14_0677	0016874 : ligase activity			2.86
	PFE1410c PFE1410c hypothetical protein 5021107:5022597 reverse MW:58826	PFE1410c				2.39
29	K2C1_HUMAN Keratin, type II cytoskeletal 1 (Cytokeratin 1) (K1) (CK 1) (67 kDa	1346343				6.39
32	Similar to human zinc-finger protein	27669054	no match	no match	no match	2.89

Table 4.7 List of proteins identified using MS from spots of interest shown on Figure 4.6 (composite CSF gel of CM compared to EN) Proteins where identified ESI-MS/MS. Red indicates spot was found in CM only and blue EN only.

Spot No	Description	Accession	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage (%)
7	similar to KIAA0456 protein [imported] - human [Rattus norvegicus]	27711750	no match	no match	no match	1.62
18	Chain A, Transthyretin (Also Called Prealbumin)	494652	0005515 : protein binding; 0005496 : steroid binding; 0019840 : isoprenoid binding; 0019842 : vitamin binding; 0005215 : transporter activity	0005576 : extracellular region	0006810 : transport; 0006575 : amino acid derivative metabolism; 0009308 : amine metabolism; 0042445 : hormone metabolism	21.26
23	K2C1_HUMAN Keratin, type II cytoskeletal 1 (Cytokeratin 1) (K1) (CK 1) (67 kDa	1346343	0030246 : carbohydrate binding; 0004871 : signal transducer activity; 0005198 : structural molecule activity; 0005515 : protein binding	0016020 : membrane; 0005856 : cytoskeleton	0050789 : regulation of biological process; 0050817 : coagulation; 0050896 : response to stimulus; 0006800 : oxygen and reactive oxygen species metabolism; 0007275 : development; 0019538 : protein metabolism	6.39

Table 4.8 List of proteins identified using MS from spots of interest shown on Figure 4.7 Gel A plasma (CM compared to ABM) Proteins where identified using ESI-MS/MS. Green indicates an increase in expression by a factor of 2 or more, yellow a decrease by a factor of 2 or more and blue spots were unique to CM

Spot No	Accession Number	PROTEIN NAME	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage%
1	4505529	orosomucoid 2; alpha-1-acid glycoprotein, type 2 [Homo sapiens]	0005488 : binding	0005576 : extracellular region	0050896 : response to stimulus	11.94
	9257232	orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1); a	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	10.45
2	-					
3	4507143	Sorting nexin 3	0008289 : lipid binding; 0005515 : protein binding		0007154 : cell communication	32
	3126979	Sorting nexin-3 (Protein SDP3)	0008289 : lipid binding; 0005515 : protein binding	0005737 : cytoplasm	0015031 : protein transport; 0007154 : cell communication; 0006810 : transport; 0016044 : membrane organization and biogenesis; 0016192 : vesicle-mediated transport	32
4	1703025	A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	16.75
	4507871	vav 2 oncogene; Oncogene VAV2 [Homo sapiens]	no match			4.21
	15990507	Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1- antiproteinase)	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	25
	110350939	Alpha-1 antitrypsin variant	0030234 : enzyme regulator activity			25
	177836	Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1- antiproteinase)	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	25
5	-					
6	7705348	M-phase phosphoprotein 1; mitotic kinesin-like protein [Homo sapiens]	no match			2.42
7	-					
8	-					
9	-					
10	-					

Table 4.9 List of proteins identified using MS from spots of interest shown on Figure 4.7 Gel B plasma (CM compared to EN). Proteins where identified using ESI-MS/MS. Green indicates an increase in expression by a factor of 2 or more, yellow a decrease by a factor of 2 or more and blue spots were unique to CM

Spot No	Accession Number	PROTEIN NAME	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage %
1	4505529	orosomucoid 2; alpha-1-acid glycoprotein, type 2 [Homo sapiens]	0005488 : binding	0005576 : extracellular region	0050896 : response to stimulus	11.94
	9257232	orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1); a	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	10.45
2	4505529	orosomucoid 2; alpha-1-acid glycoprotein, type 2 [Homo sapiens]	0005488 : binding	0005576 : extracellular region	0050896 : response to stimulus	11.94
	9257232	orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1); a	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	15.42
3	-					
4	-					
5	-					
6	-					
7	-					
8	4507143	Sorting nexin-3 (Protein SDP3)	0008289 : lipid binding; 0005515 : protein binding	0005737 : cytoplasm	0015031 : protein transport; 0007154 : cell communication; 0006810 : transport; 0016044 : membrane organization and biogenesis; 0016192 : vesicle-mediated transport	32
9	1703025	A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	16.75
	4507871	vav 2 oncogene; Oncogene VAV2 [Homo sapiens]	no match			4.21
	15990507	Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1-antiproteinase)	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	25
	110350939	Alpha-1 antitrypsin variant	0030234 : enzyme regulator activity			25
	177836	Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1-antiproteinase)	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	25
10	-					

Table 4.9 continued

Spot No	Accession Number	PROTEIN NAME	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage %
11	-					
12	2914175	A Chain A, Crystal Structure Of Human Apolipoprotein A-I	0005215 : transporter activity; 0005515 : protein binding; 0008289 : lipid binding	0005576 : extracellular region	0006869 : lipid transport; 0019538 : protein metabolism; 0050789 : regulation of biological process; 0006629 : lipid metabolism; 0006810 : transport; 0008150 : biological_process; 0006066 : alcohol metabolism	18.41
	235865	sperm activating protein subunit I, apolipoprotein A	no match			50
	90108664	Apolipoprotein A-I precursor (Apo-AI) (ApoA-I) [Contains: Apolipoprotein A-I(1-242)]	0005215 : transporter activity; 0005515 : protein binding; 0008289 : lipid binding	0005576 : extracellular region	0006869 : lipid transport; 0019538 : protein metabolism; 0050789 : regulation of biological process; 0006629 : lipid metabolism; 0006810 : transport; 0008150 : biological_process; 0006066 : alcohol metabolism	40
13	-					
14	-					

Table 4.10 Protein identified using MS from spots of interest shown on Figure 4.8 Gel A CSF (CM compared to ABM). Proteins where identified using ESI-MS/MS. Green indicates an increase in expression by a factor of 2 or more, yellow a decrease by a factor of 2 or more and blue spots were unique to CM

Spot No	Accession Number	PROTEIN NAME	GO SLIM Molecular Function	GO SLIM Cellular Component	GO SLIM Biological Process	Sequence Coverage %
1	-					
2	-					
3	PF10_0224	PF10_0224 PF10_0224 dynein heavy chain, putative 12042889:12059952 reverse MW:671885	0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0046907 : intracellular transport	2.25
	6912728	transmembrane tryptase preproprotein; mast cell tryptase; mast cell pro	No match	No match	No match	4.79
	4757764	ras homolog gene family, member B; RhoB; Aplysia RAS-related homolog 6	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0005515 : protein binding	0016020 : membrane; 0005622 : intracellular; 0005768 : endosome; 0005634 : nucleus	0015031 : protein transport; 0007275 : development; 0008219 : cell death; 0007165 : signal transduction; 0007155 : cell adhesion; 0006996 : organelle organization and biogenesis; 0016192 : vesicle-mediated transport; 0046907 : intracellular transport; 0006810 : transport; 0007049 : cell cycle; 0050789 : regulation of biological process	7.14
4	-					
5	-					
6	-					
7	-					
8						
9						
10						

4.4 Discussion

The work described in this chapter demonstrated the use of 2-DE gel analysis to study unique LAPS associated with CM, the most severe neurological complication of falciparum malaria that leads to significant mortality and morbidity including neurological sequelae. *P. falciparum* appears to have a particular propensity for the brain, since the other human malarial parasites rarely cause neurological dysfunction. This is thought to be directly linked to sequestration and binding of the infected red cell within the brain microvasculature (Gitau and Newton, 2005). The working hypothesis was that by studying both the plasma and CSF proteome it would be possible to identify unique biomarkers for the early-stage diagnosis or the staging of the neurological disease, offer potential insights into the biochemical perturbations induced by sequestration including affects on specific neuronal populations, and clarify the basic molecular basis of the pathology.

Patients in this study had a mean age of 34, 26, and 37 months for CM, ABM, and EN respectively. The number of seizures per disease group ranged between 8 and 11 seizures and none of the patients were severely anemic. The average mortality was 25% , 17%, 33% for CM, ABM and EN.

The functional integrity of the BBB is usually assessed by examining the passage of molecules into the CSF. The passage of the molecules from the blood into the brain is determined by the size, and charge of the molecules, and the presence of specific receptors in the BBB for transporting molecules. Thus in human studies, the ratio of CSF to blood concentrations are often used to determine the integrity of the BBB. In particular, the concentrations of relatively small proteins such as albumin are used widely. An elevated CSF albumin/serum albumin ratio indicates increased permeability of the BBB. The three disease groups studied are all infectious diseases associated with increases in cytokines and acute phase proteins which are common to these diseases. However, there are fundamental differences as the BBB is greatly impaired in ABM, less so in CM and least in EN. The

initial protein content of CSF determined using the Bradford method (table 4.4) gives a clear indication that protein levels increase with BBB impairment (3.34g/L, 0.53g/L and 0.23g/L for ABM, CM and EN respectively). Normal levels of total protein levels of CSF in children remain unknown as it is impossible to obtain.

Majority of the proteins differentially expressed are acute phase proteins (APPs) which either decrease (ne gative) or increase (positive) during an acute phase response. The functions of APPs are many; some positive APPs participate in host-adaptive and host-defense mechanisms by binding to foreign substances or by having opsonizing activities and modulating phagocytic cell functions. Other APPs have more specific actions, such as inhibiting serine proteinases or serving as transport proteins with antioxidant activity. On the other hand, negative APPs such as transthyretin, retinol binding protein, transferrin, and albumin have no apparent immune function (Schreiber et al., 1989). Their main role is to transport nutrients and, therefore, their reduction during infection and inflammation may lower the concentration of specific nutrients such as vitamin A in malarial infection (Schreiber et al., 1989).

At present, 2-DE is a popular method to separate a complex mixture of proteins. The ability to monitor many proteins simultaneously yields a global view of protein expression and post-translational modification, which is much more informative than monitoring a few proteins (Powell and Timperman, 2004). However, limitations of 2-DE, including reproducibility and limited sample load continue to create problems when trying to analyse complex samples such as plasma and CSF. The application of a commercial immobilized pH gradient (IPG) strip has greatly improved 2-DE reproducibility and although not applied in this study the use of longer (18 or 24 cm) and narrower-p*I* ranges strips have increased significantly the electrophoretic resolution of proteins and the sample load (Yuan and Desiderio, 2005b). Some 2-DE systems can simultaneously analyse up to 12 gels to minimise any difference between-gels that would

be caused by instrumental conditions. To account for any technical and biological variations, 12 patient samples per disease group were analysed in duplicate in thus study.

For quantitative analysis the samples were stained with silver stain. Although silver stain is not very compatible with MS, it was not possible to identify many proteins after staining the CSF gels with coomassie due to the low amount of protein found in the sample and the limited sensitivity of coomassie staining. Most proteomic studies of neurological pathologies that use CSF try to overcome the problem of low protein content by pooling samples (Fukuyama et al., 2000, Puchades et al., 2003, Davidsson et al., 2002a, Davidsson et al., 2002b, Yuan et al., 2002, Hammack et al., 2003). However, in proteomics, one problem with pooling is how to sort out highly abundant single case contaminations from the pool (low abundant contaminations are diluted). This can be achieved by differential displays from the total pools (e.g. each 30 patients) and a set of subpools (each group 5 × 6 patients). In this case a minimum of six differential displays can unambiguously identify the general serum biomarkers and discard single fates only apparent in one of the subpools (Schrattenholz, 2004). Another major disadvantage of pooling samples before proteomics analysis is that information on the sensitivity and reliability of each candidate marker in detecting disease onset and/or progression will be lost, as this information can only come from analyzing sets of individual samples (Jing, 2007). In this study samples were not pooled and any proteins identified were found in majority of the patients in the disease phenotype.

The overall success of differential protein display in proteome research depends critically on the accuracy and the reliability of the analysis software. In addition, the software has a profound effect on the interpretation of the results obtained, and the amount of user intervention demanded during the analysis. In this study both PDQuest® and Progenesis® software were used to analyse the gels. While PDQuest gave a robust spot overlay analysis, when matched spots were missing in some gels, the PDQuest software assigned an arbitrary value to the missing spots, introducing a false normalised quantity in the calculation of the

Student's *t*-test and for this reason all quantitation was done using the Progenesis® software. However as our results showed spots can contain more than one protein due to insufficient resolution. In such cases the measured spot volume ratios will not reflect the real quantitative situation, a good example being spot 10 (**Figure 4.5**) which had more than one protein attributed to it. Using Progenesis® Alpha-1-antitrypsin precursor and it's variant were found to be unique to CM plasma (Spot 4, **Table 4.5** and spot 9, **Table 4.6**). However, this protein was also identified on a different spot (spot 7 and 8 Table 4.1) but only spot 8 was identified as unique using PDQuest®. The identification of different spots as the same protein would indicate that different isoforms of this protein migrate differently and understanding what the roles the different isoforms may have in a disease would be very useful. The result on alpha-1- antitrypsin justifies the need to understand how it's different isoforms work especially because one would expect to see it in all the disease groups studied as the protein is normally increased in acute inflammation. Chain A transthyretin (retinol binding protein) was found uniquely expressed in CSF from CM (Spot 18, Table **4.4**). This protein has sometimes been attributed to transfer of thyroxine from plasma to the brain (Dickson et al., 1987, Schreiber et al., 1990, Chanoine et al., 1992, Southwell et al., 1993). As discussed in chapter 3 this acute phase protein is decreased in an acute phase response and the presence of it in CM would suggest that the response is not as strong as in the other phenotypes

Orosomucoid 2; alpha-1-acid glycoprotein, type 2 is one of the proteins that was found to be up-regulated in plasma CM when compared with both EN and ABM gels. This protein has been shown to inhibit neutrophil activation and production of superoxide radicals and TNF-α (Costello et al., 1984, Scuderi et al., 1989).49–51 Furthermore, increased production of orosmucoid, caeruloplasmin, and of glutathione, enhance antioxidant defences and limit the stimulatory effects of oxidant molecules on cytokine production. The sorting nexin (SNX) protein family is implicated in regulating membrane traffic, and SNX3 is associated with endosomal function. In our study SNX3 was expressed in lower content in CM compared to both ABM and EN (spot 3 **Table 4.5** and spot 8 **Table 4.6**).

Two different types of mass spectrometry were used in this study. MALDI-ToF mass spectrometry is the widely used approach in proteomics for identification of proteins using PMF data from spots. During this study about 60% of proteins identified from coomassie stained spots were identified using MALDI-ToF and about 97% of the proteins identified using LC-MS/MS. A study comparing the 2 techniques found that identification of 2-DE separated proteins that matched with the most number of peptides by LC-MS/MS were also the proteins exclusively identified as the number 1 ranked proteins in the first round of a PMF search. No unique proteins were identified exclusively by PMF (Lim et al., 2003). This study shows that MALDI-ToF is complementary to MS/MS identification of 2-DE proteins, but lacks the dynamic range of LC-MS/MS. Improved dynamic range is obtained through the LC separation process to enable the acquisition of tandem mass spectra for proteins present in a spot in less abundance. MS/MS database searching uses fragmentation patterns indicative of specific amino acid sequences to match a protein and thus has a higher level of specificity than PMF. Thus, only one peptide MS/MS spectrum with good signal to noise, fragmentation, and length can identify a protein in the database. Another drawback of the PMF approach in protein identification can be found when judging a search result. If the score of the first ranked protein is around or below the boundary of the criteria set by the software, the user cannot be confident with the result and has to manually inspect the data to obtain additional information (e.g., sequence) to increase the confidence of the result. In most cases, the protein has to be reanalyzed preferably by MS/MS

Although 2-DE is a powerful technique, one of its limitations is that it remains relatively low throughput and requires large amounts of starting material (~50μg) with low sensitivity for detection of low abundance proteins such as cytokines and signalling molecules. In addition, certain basic proteins, and very high- or very low-molecular weight proteins are not separated well by 2-DE. Techniques such as free-flow electrophoresis (FFE) have been developed to help resolve complex protein mixtures using a combination of FFE (liquid based IEF method) and 2-DE (Hoffmann et al., 2001). The use of narrow range, overlapping pH gradients in the

first dimension also improves the number of proteins visualised on 2D gels (Gorg et al., 2004). Until recently, a limitation in 2-DE technology was the reproducibility, necessitating the use of multiple gels to obtain statistical validity. A major advance in this area has come from the introduction of Cy dye fluorophores for pre-labelling of protein samples Two-dimensional fluorescence difference gel electrophoresis (2D-DIGE) technology adds a quantitative component to conventional 2-DE analyses, allowing for comparison of protein expression changes across multiple samples simultaneously without gel-to-gel variation, and hence with statistical confidence (Tonge et al., 2001, Unlu et al., 1997). 2D-DIGE utilises Cy dye fluorophores for protein labelling prior to fractionation. This facilitates multiplexing of protein samples, allowing for direct comparison of different samples within the one gel (Unlu et al., 1997), and more importantly, enables the introduction of a standardised internal control.

Separation of a very complex mixture of larger proteins is best performed by 2-DE. While the presence of large proteins hinders the separation of smaller ones, the presence of small proteins does not interfere with 2-DE, so they do not need to be removed if one is only interested in the larger ones. The advantages of 2-DE are resolving power sufficient often to separate various forms of the same proteins (e.g. proteins with different glycosylation or phosphorylation patterns), multiplexing (hundreds or thousands of proteins separated and visualised per gel), and ease of comparison and quantitation (especially through the technique of difference gel electrophoresis known as DIGE). The disadvantages include marginal applicability to proteins with extreme isoelectric points or hydrophobicity (membrane proteins) and to proteins of small size, and technical difficulties associated with the technique. For less complex samples, digestion prior to separation avoids gels and allows more convenient analysis but, since not all fragments are identified, this approach forfeits the opportunity offered by gel-based analysis to obtain more information on the different (post-translationally modified) forms of the same gene.

Some of the main challenges facing 2-DE of plasma and CSF include the great dynamic range of protein abundance and a wide range of protein properties, including mass, isoelectric point, extent of hydrophobicity and post-translational modifications. One strategy to simplify complex samples is to deplete abundant proteins with a variety of affinity-based depletion columns, *e.g.*, a multiple affinity removal system by Agilent Technologies, which is commercially available (Shen et al., 2006). This technology has been used for various body fluids, including serum, saliva, and CSF (Ramstrom et al., 2005, Li and Lee, 2004, Bjorhall et al., 2005), resulting in more proteins being identified after depletion of abundant proteins. However, as discussed in chapter 3, others have presented data questioning the utility of this approach (Omenn et al., 2005). In addition, as affinity depletion of abundant proteins is performed at non-denaturing conditions, it is likely that a significant number of other proteins will be depleted along with abundant proteins. This is particularly true for albumin, as it is one of the major carrier proteins in all body fluids. Given the caveats associated with affinity depletion of abundant proteins, making use of columns with various binding mechanisms, *e.g.*, cation exchange, ionic exchange, and solid phase extraction, may better isolate protein populations, hoping that CSF proteins can be fractionated and thereby enriching proteins of low abundance.

Various studies have applied a straightforward method, graduating organic fractionation approach, which is comparable to both qualitative and quantitative proteomics for simplifying CSF protein profiles (Abdi et al., 2006, Zhang et al., 2005a, Zhang et al., 2005b). Briefly, CSF was mixed with 1.5 volume of ACN to generate the first pellet (P1), and then the supernatant was mixed with another 1.5 volume of ACN (*i.e.*, a final 3.0 volume of ACN) to generate the second pellet (P2) and a supernatant (S2), which was dialyzed with a porous (0.5 kDa) membrane to desalt. With this approach, more than 90% of albumin and IgGs were found in the first pellet (Zhang et al., 2005a). However, increasing another step prior to 2-DE would further increase the time taken and hence a non-gel based method of separation is best explored for high throughput proteomics.

4.5 Conclusions

Undoubtedly, various non-gel-based schemes that rely on liquid based separations of proteins or peptides, with or without tagging, will have utility for disease proteomics, particularly given their potential for automation. Additionally, advances in microfluidic technology will likely allow automated separation of proteins in complex lysates using much reduced sample amounts. Microfluidic systems already have been integrated with mass spectrometry for protein digestion and identification (Brivio et al., 2002). In this chapter, it has been shown that by applying the 2-DE technique protein profiles of severe disease can be created. Although some differences in the protein profiles of the different disease groups were identified, the limitations of the technique did not help identify as many proteins as expected. The next chapter describes a non-gel based method of protein separation prior to mass spectrometry and explores the utility of the tool.

5 Modified 'Shotgun' Analysis of Plasma and CSF Proteomes from Cerebral Malaria patients

5.1 Introduction

The use of the 2-DE approach for global proteomics analysis of plasma and CSF from children diagnosed with cerebral malaria was discussed in chapter 4. While this approach continues to be the mainstay in proteomic analysis, there are well known limitations to the technique, especially for complex samples such as plasma and CSF. As discussed in chapter 4, one of the limitations is the fact that the sample load capacity of 2-DE gels is severely limited by the presence of high abundance proteins which makes detection of potential biomarker proteins found in relatively low concentrations very challenging. Moreover, as discussed in chapter 2, strategies for depletion of high abundance proteins, although very useful, have the main disadvantage of binding a large number of proteins non-specifically resulting in their loss to analysis. One alternative to 2-DE is the use of non-gel based separation techniques such as liquid chromatography.

Chromatographic techniques have developed into powerful separation techniques, capable of separating large numbers of proteins and peptides. However, any given chromatographic technique will still only separate a small fraction of proteins in a complex mixture. The driving analytical platform behind the emergence of non 2-DE MS-based proteomic methods has been the coupling of various orthogonal chromatographic separations prior to mass spectrometry. This has led to the introduction of what is known as 'shotgun proteomics' which refers to the direct analysis of complex protein mixtures to generate a global profile of the protein complement within the mixture (Wu and MacCoss, 2002).

Initial shotgun approaches involved digestion of the complex sample followed by multidimensional separation of the peptide mixture prior to mass spectrometry. This method is commonly referred to as the "bottom-up method" (Kettman et al., 2001), and is the basis of the MudPit Technology (Link et al., 1999). It uses two dimensional chromatography where the first dimension separates the proteins according to charge using a strong cation exchange (SCX) column. The advantage of this method is that breaking proteins into peptides allows separation and identification of those proteins that are missed by gel electrophoresis, such as hydrophobic, membrane bound proteins and LAPs. However, one of the main disadvantages of this approach is the complexity of the peptide mixture (Washburn et al., 2001, Wolters et al., 2001), resulting in ion suppression in ESI (Sterner et al., 2000) which hampers the subsequent protein identification. This chapter describes a pre-fractionation technique prior to protein digestion that can be used to overcome this limitation. It involves using a monolithic column and a LC-MS method that has a shorter gradient time than the usual reversed phase (RP) - MS.

5.2 Materials and Methods

5.2.1 S ample Source

As described in Chapter 4, plasma and CSF samples collected from children in Kilifi with a diagnosis of cerebral malaria (CM) (n=12) were compared to CSF and plasma samples collected from children in the same population with a diagnosis of non-specific encephalopathy (EN) (n=12) and acute bacterial meningitis (ABM) (n=12).

Table 5.1 Clinical Characteristics of Study Patients

	Cerebral Malaria (CM) (n=12)	Acute Bacterial Meningitis (ABM) (n=12)	Undiagnosed Encephalopathy (EN) (n=12)
Age[7]	*33.89 (21.75)*	*52.41 (51.33)*	*24.85 (15.64)*
Sex (%Male)	*33*	*42*	*33*
Duration of illness before admission to hospital[8]	*2.5 (1.31)*	*2 (1.18)*	*2*
Number with Seizures	*8*	*8*	*10*
Number with Coma	*11*	*5*	*8*
Hb (s.d.)	*6.89 (1.94)*	*8.6 (2.4)*	*10 (1.05)*
Duration in Hospital[9]	*4.4 (2.6)*	*34 (54.50)*	*4.7 (3.04)*
Mortality (%)	*0*	*33*	*17*

[7] Average age in months (s.d)
[8] Average number of days (s.d)
[9] Average number of days (s.d.)

5.2.2 Fi rst dimension Liquid Chromatography

The sample preparation protocol described previously chapter 2 section 2.2.3.2 was followed for the fractionation of proteins in CSF and plasma. Samples were first separated as described in Chapter 2, section 2.2.3.2. To normalise for protein content, an equivalent of 100µg of CSF protein and 200µg of plasma protein was injected onto the column. Protein content had been estimated as describe in section 2.1.2.1. Fractions were collected every 30 seconds starting from 1 minute to 25 minutes. The fractions were then dried down overnight in an oven set at 50°C. 25µl of 100mM ammonium bicarbonate (ABC) was added to the sample using a repeat pipette followed by 5µl of a 20µg/mL solution of trypsin in 25mM ABC. Samples were thoroughly mixed using a multi-channel pipette and incubated overnight at 37°C to achieve complete digestion.

5.2.3 S econd Dimension Reversed Phase Liquid Chromatography

The second dimension HPLC separation was performed on an UltiMate™ 3000 LC system (Dionex UK). An aliquot (5 µL) of the sample solution (in a 10 µL sample loop) was injected from the autosampler (using the full loop mode) onto a Monolithic Capillary Column (200 µm i.d. x 5 cm; Dionex Corporation). Solvent A was 2.5% v/v Acetonitrile (ACN) in water with 0.1% formic acid and solvent B was 90% v/v ACN in water with 0.1% formic acid. The flow rate on the micro pump in the LC system was maintained at 1.5 µL/min. The gradient was started at 5% solvent B, ramped to 40% solvent B in 12 min, and then to 90% solvent B for another 2 min. Solvent B was then decreased to 5% and this was maintained to the end of the run at 27 minutes.

5.2.4 M ass Spectrometry

The LCQ Deca XP Plus ion trap mass spectrometer (ThermoFinnigan, USA) was equipped with a nanospray source using a PicoTip connected to the column. To optimise the spray, the tip was moved close to the entrance of the mass spectrometer (approx. 2mm) and the

spray voltage adjusted (normally 1.5kV-1.8kv) to maintain a consistent spray. The temperature of the ion transfer tube was set at 240 °C and the normalised collision energies were set at 35% for MS/MS. The mass spectrometer was operated on a data-dependent "big-three" mode where the three most intense ions in the full scan were subjected to MS/MS. Dynamic exclusion was set at a repeat count of 2, a repeat duration of 0.5 minute, and exclusion duration of 3 min. to obtain MS/MS spectra from any co-eluting peptides.

5.2.5 P rotein Identification

As described in Chapter 2 section 2.3.5.4,the peptide fragmentation data were searched against the non redundant NCBI database and the *P. falciparum* database downloaded from the Sanger Institute using the TurboSEQUEST® algorithm in BioWorks v 3.1 software provided by ThermoFinnigan™ .

5.2.6 Data base Search Parameters

All spectra were searched against the *human* and *P. falciparum* subset of the NCBInr database and *P. falciparum* database downloaded from the Sanger Institute. Searches were then repeated on the entire NCBInr database and only proteins which were called after both searches were accepted.

5.2.6.1 Protein Identifications

Proteins were accepted as legitimate identifications if they had two or more peptides with cross-correlation scores (Xcorr) of >1.5 for singly charged, >2.00 for doubly charged and >2.50 for triply charged peptides each with a delta correlation (DelCn) > 0.1.

5.2.6.2 Characterisation and Cataloguing Of Identified Proteins

Protein characterisation and cataloguing was done according to GO annotations using the PIR batch retrieval tool as described in chapter 2 section 2.3.5.5

5.3 Results

5.3.1 P rotein Identification by Mass Spectrometry

Appendices I-IV give a complete list of all identified proteins identified from the plasma and CSF. In plasma, a total of 236 *human* proteins and 323 *P. falciparum* proteins were identified from the database searches. From the 236 *human* proteins, 33 were present in CM, 40 in EN and 217 in ABM. 3 proteins were found to be unique to CM, 15 to EN and 180 to ABM. 1 protein was common to CM and EN, 13 proteins were common to CM and ABM, 8 common to EN and ABM and 16 proteins were found in all three disease groups. Of the 323 *P. falciparum* proteins identified, 160 were present in CM, 21 in EN and 237 in ABM. 79 were found to be unique to CM, 4 unique to EN and 154 unique to ABM. 66 of the *P. falciparum* proteins were common to CM and ABM and 2 were common to EN and ABM. 15 proteins were found to be common to all three disease groups. Of the 323 *P. falciparum* proteins 222 were identified as hypothetical. **Table 5.2** gives a list of the non-hypothetical unique proteins found in plasma drawn from patients with CM.

In CSF, a total of 143 human proteins were identified. Of the 143, 21 were found in CM, 34 in EN and 121 in ABM. 8 proteins were unique to CM, 11 to EN and 93 unique to ABM. There were 5 proteins found only in CM and ABM and 26 proteins found only in EN and ABM. 8 of the proteins were found in all 3 disease groups. A total of 66 *P. falciparum* proteins were identified and of these, 7 were found in CM, 3 in EN and 61 in ABM. 2 of the proteins were unique to CM, 3 unique to EN and 56 unique to ABM. 5 of

the proteins were found common to both CM and ABM. 41 human and 34 *P. falciparum* proteins were found in both CSF and plasma and these are summarised in **Appendices V and VI**. **Table 5.2** gives a list of the unique proteins found in CSF drawn from children diagnosed with CM. The 2 *P. falciparum* proteins were not included in the list as they were hypothetical proteins. A list of proteins found in both plasma and CSF drawn from children with CM are listed on **Table 5.3.** **Appendix V** gives a complete list of the proteins found in both plasma and CSF for all disease groups.

Table 5.2 List of Proteins Found only in Plasma Drawn From Children with CM

Reference	Accession	Peptides	% Peptide Coverage	GO Slim Function	GO Slim Cellular Component	GO Slim Biological Process	Organism Name
Orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1 ; a	9257232	5	12.24	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus	*Homo Sapiens*
Chain B, Crystal Structure Of A Human Fcg Receptor In Complex With An Fc Fragment Of Igg1	14277820	9	12.47	0005515 : protein binding; 0003823 : antigen binding	0000267 : cell fraction	0050896 : response to stimulus	*Homo Sapiens*
MAL13P1.134 MAL13P1.134 helicase, putative 18100324:18104807 reverse MW:130703	MAL13P1.134	2	8.21	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0003824 : catalytic activity	0005634 : nucleus	0006139 : nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism	*Plasmodium falciparum*
MAL13P1.221 atcasE aspartate carbamoyltransferase 18869475:18870910 forward MW:43252	MAL13P1.221	2	15.41	0043176 : amine binding; 0016740 : transferase activity		0006520 : amino acid metabolism; 0009117 : nucleotide metabolism; 0009112 : nucleobase metabolism	*Plasmodium falciparum*
MAL7P1.228 HSP70 Heat Shock 70 KDa Protein, (HSP70) 6722615:6724917 forward MW:73214	MAL7P1.228	3	14.23				*Plasmodium falciparum*

Table 5.2 cont'd

Reference	Accession	Peptides	% Peptide Coverage	GO Slim Function	GO Slim Cellular Component	GO Slim Biological Process	Organism Name
PF08_0141 VAR erythrocyte membrane protein 1 (PfEMP1) 8148816:8158149 reverse MW:330699	PF08_0141	2	6.53	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis	*Plasmodium falciparum*
PF10_0165 PF10_0165 DNA polymerase delta catalytic subunit 11764971:11768255 forward MW:126885	PF10_0165	15	12.91	0016740 : transferase activity; 0003676 : nucleic acid binding; 0016787 : hydrolase activity; 0000166 : nucleotide binding	0005634 : nucleus	0006260 : DNA replication	*Plasmodium falciparum*
PF11_0270 PF11_0270 threonine – tRNA ligase, putative 13781382:13784423 reverse MW:119546	PF11_0270	2	12.31	0016874 : ligase activity; 0000166 : nucleotide binding		0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism	*Plasmodium falciparum*
PF14_0102 PF14_0102 rhoptry-associated protein 1 20392910:20395258 forward MW:90053	PF14_0102	2	11.24				*Plasmodium falciparum*
PF14_0202 PF14_0202 dynein-associated protein, putative 20836037:20837234 reverse MW:30842	PF14_0202	6	7.24				*Plasmodium falciparum*
PF14_0281 PF14_0281 aspartyl protease, putative 21160759:21163876 forward MW:74183	PF14_0281	2	6.23	0016787 : hydrolase activity		0019538 : protein metabolism	*Plasmodium falciparum*

Table 5.2 cont'd

Reference	Accession	Peptides	% Peptide Coverage	GO Slim Function	GO Slim Cellular Component	GO Slim Biological Process	Organism Name
PF14_0334 PF14_0334 NAD(P)H-dependent glutamate synthase, putative 21394312:21403632 reverse MW:353	PF14_0334	2	11.22	0003824 : catalytic activity; 0043167 : ion binding; 0016491 : oxidoreductase activity	0016020 : membrane	0006091 : generation of precursor metabolites and energy; 0008152 : metabolism; 0006520 : amino acid metabolism; 0006807 : nitrogen compound metabolism	*Plasmodium falciparum*
PF14_0364 PF14_0364 cleavage and polyadenylation specifity factor protein, putative 21529938:215325	PF14_0364	139	30.34	0016787 : hydrolase activity			*Plasmodium falciparum*
PFA0345w PFA0345w centrin, putative 293860:295186 forward MW:19599	PFA0345w	2	6.42				*Plasmodium falciparum*
PFB0010w PFB0010w erythrocyte membrane protein 1 (PfEMP1) 668524:674460 forward MW:196552	PFB0010w	2	8.21	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis	*Plasmodium falciparum*
PFB0405w PFB0405w transmission-blocking target antigen s230 precursor 1013730:1023137 forward MW:36	PFB0405w	5	5.71		0016020 : membrane		*Plasmodium falciparum*
PFB0665w PFB0665w Ser/Thr protein kinase, putative 1237488:1242632 forward MW:204291	PFB0665w	2	7.21	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism	*Plasmodium falciparum*
PFC0805w PFC0805w DNA-directed RNA polymerase II, putative 2335530:2342903 forward MW:278679	PFC0805w	2	5.43				*Plasmodium falciparum*
PFC0870w PFC0870w elongation factor 1 (EF-1), putative 2404977:2405759 forward MW:17706	PFC0870w	3	5.87				*Plasmodium falciparum*

Table 5.2 cont'd

Reference	Accession	Peptides	% Peptide Coverage	GO Slim Function	GO Slim Cellular Component	GO Slim Biological Process	Organism Name
PFD0685c PFD0685c chromosome associated protein, putative 3284355:3290588 reverse MW:141227	PFD0685c	5	16.71	0000166 : nucleotide binding; 0005515 : protein binding; 0016787 : hydrolase activity	0016020 : membrane; 0005694 : chromosome	0006259 : DNA metabolism; 0006323 : DNA packaging	*Plasmodium falciparum*
PFD0790c PFD0790c DNA replication licensing factor, putative 3368556:3372953 reverse MW:171315	PFD0790c	3	8.41	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding		0006260 : DNA replication	*Plasmodium falciparum*
PFD0995c VAR erythrocyte membrane protein 1 (PfEMP1) 3589951:3596795 reverse MW:247612	PFD0995c	2	9.43	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis	*Plasmodium falciparum*
PFD1045c PFD1045c erythrocyte membrane-associated antigen, putative 3659383:3672168 reverse MW:5085	PFD1045c	2	5.43				*Plasmodium falciparum*
PFD1050w PFD1050w alpha-tubulin ii 3677408:3679087 forward MW:49691	PFD1050w	6	8.41	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0005198 : structural molecule activity	0043234 : protein complex; 0005856 : cytoskeleton	0019538 : protein metabolism; 0016043 : cell organization and biogenesis; 0046907 : intracellular transport	*Plasmodium falciparum*
PFE1640w VAR-like erythrocyte membrane protein 1 (PfEMP1), truncated 5188058:5197552 forward MW:367	PFE1640w	2	11.24	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis	*Plasmodium falciparum*

Table 5.2 cont'd

Reference	Accession	Peptides	% Peptide Coverage	GO Slim Function	GO Slim Cellular Component	GO Slim Biological Process	Organism Name
PFF1140c PFF1140c ATP-dependent DEAD box helicase, putative 6157278:6160691 reverse MW:136287	PFF1140c	3	6.32	0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity			*Plasmodium falciparum*
PFF1430c PFF1430c transmembrane amino acid transporter protein, putative 6412095:6414152 reverse MW	PFF1430c	2	6.84		0016020 : membrane		*Plasmodium falciparum*
PFF1450w PFF1450w sec14-like cytosolic factor or phosphatidylinositol/phosphatidylcholine transfer	PFF1450w	5	10.31				*Plasmodium falciparum*
PFL0770w 2277.t00155 seryl-tRNA synthetase, putative 15449918:15451774 forward MW:73273	PFL0770w	2	15.21	0016874 : ligase activity; 0000166 : nucleotide binding		0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism	*Plasmodium falciparum*
PFL0785c 2277.t00158 signal recognition particle 19 kd protein, putative 15456777:15457235 reverse	PFL0785c	7	13.20	0003676 : nucleic acid binding	0005737 : cytoplasm; 0030529 : ribonucleoprotein complex	0015031 : protein transport; 0046907 : intracellular transport	*Plasmodium falciparum*
PFL1950w VAR erythrocyte membrane protein 1 (PfEMP1) 16499524:16508461 forward MW:300289	PFL1950w	4	12.20	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis	*Plasmodium falciparum*

Table 5.3 List of Proteins Found only in CSF Drawn From Children with CM

Reference	Accession	Peptides	% Peptide coverage	GO Slim Biological Function	GO Slim Cellular Component.	GO Slim Biological Process	Organism
guanylate cyclase 2F; RetGC-2; guanylate cyclase 2D-like, membrane (ret	4504219	6	4.78				*Homo Sapiens*
secretogranin II precursor; Chromogranin C (secre	4506801	3	5.67				*Homo Sapiens*
solute carrier family 5 (sodium/glucose cotransporter), member 1; Human	4507031	2	8.4	0043167 : ion binding; 0005215 : transporter activity	0016020 : membrane; 0045177 : apical part of cell; 0042995 : cell projection	0008150 : biological_process; 0008643 : carbohydrate transport; 0006811 : ion transport; 0006810 : transport; 0007275 : development	*Homo Sapiens*
alanine-glyoxylate aminotransferase; alanine-glyoxylate aminotransferas	4557289	3	8.93	0016740 : transferase activity; 0005515 : protein binding; 0019842 : vitamin binding; 0048037 : cofactor binding	0005739 : mitochondrion; 0042579 : microbody	0008152 : metabolism; 0006081 : aldehyde metabolism; 0006082 : organic acid metabolism; 0006996 : organelle organization and biogenesis; 0015031 : protein transport; 0046907 : intracellular transport	*Homo Sapiens*
keratin 2a [*Homo Sapiens*]	4557703	4	6.82				*Homo Sapiens*
phospholipase C, beta 2 [*Homo Sapiens*]	4758938	2	2.12				*Homo Sapiens*
endothelial lipase precursor; endothelial cell-derived lipase [Homo sap	5174497	3	5.20	0030246 : carbohydrate binding; 0016787 : hydrolase activity; 0003824 : catalytic activity; 0005515 : protein binding		0006629 : lipid metabolism	*Homo Sapiens*
SPCQ_HUMAN Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3	17368942	6	2.81	0005515 : protein binding; 0005198 : structural molecule activity	0016020 : membrane; 0005737 : cytoplasm; 0005634 : nucleus; 0043005 : neuron projection; 0005856 : cytoskeleton; 0043025 : cell soma	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0007267 : cell-cell signaling; 0016192 : vesicle-mediated transport; 0008150 : biological_process	*Homo Sapiens*

Table 5.4 List of Proteins found in both Plasma and CSF from children with CM

Reference	Accession	Plasma	CSF	GO Slim Molecular Function.	GO Slim Cellular Component	GO Slim Biological Process	Organism
B34611 3',5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35 alpha chain - human	105117	CM, EN, ABM	CM, ABM	0016787 : hydrolase activity; 0003824 : catalytic activity	0016020 : membrane	0008150 : biological process; 0007165 : signal transduction; 0050896 : response to stimulus	*Homo Sapiens*
K2C1_HUMAN Keratin, type II cytoskeletal 1 (Cytokeratin 1 (K1 (CK 1 (67 kDa	1346343	CM, ABM	CM, EN, ABM	0030246 : carbohydrate binding; 0004871 : signal transducer activity; 0005198 : structural molecule activity; 0005515 : protein binding	0016020 : membrane; 0005856 : cytoskeleton	0050789 : regulation of biological process; 0050817 : coagulation; 0050896 : response to stimulus; 0006800 : oxygen and reactive oxygen species metabolism; 0007275 : development; 0019538 : protein metabolism	*Homo Sapiens*
C Chain C, Human Serum Transferrin, Recombinant N-Terminal Lobe, Apo Form	4389232	CM, EN, ABM	CM, EN, ABM	0043167 : ion binding	0005576 : extracellular region; 0005768 : endosome; 0016023 : cytoplasmic membrane-bound vesicle	0006810 : transport; 0019725 : cell homeostasis; 0006811 : ion transport	*Homo Sapiens*

Table 5.4 cont'd

Reference	Accession	Plasma	CSF	GO Slim Molecular Function.	GO Slim Cellular Component	GO Slim Biological Process	Organism
Human Serum Albumin In A Complex With Myristic Acid And Tri-Iodobenzoic Acid	4389275	CM, EN, ABM	CM, EN, ABM	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization	*Homo Sapiens*
guanylate cyclase 2F; RetGC-2; guanylate cyclase 2D-like, membrane (ret	4504219	CM, ABM	CM				*Homo Sapiens*
v-rel reticuloendotheliosis viral oncogene homolog B, nuclear factor of	5730007	CM, ABM	CM, ABM				*Homo Sapiens*
ATP-binding cassette, sub-family B, member 10 [*Homo Sapiens*]	9961244	CM, EN, ABM	CM, EN, ABM	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane; 0005739 : mitochondrion; 0043190 : ATP-binding cassette (ABC) transporter complex	0006810 : transport	*Homo Sapiens*
SPCQ_HUMAN Spectrin beta chain, brain 3 (Spectrin, non-	17368942	CM, EN, ABM	CM	0005515 : protein binding; 0005198 : structural molecule activity	0016020 : membrane; 0005737 : cytoplasm; 0005634 : nucleus; 0043005 : neuron projection; 0005856 : cytoskeleton; 0043025 : cell soma	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0007267 : cell-cell signaling; 0016192 : vesicle-mediated transport; 0008150 : biological_process	*Homo Sapiens*

Table 5.4 cont'd

Reference	Accession	Plasma	CSF	GO Slim Molecular Function.	GO Slim Cellular Component	GO Slim Biological Process	Organism
MAL13P1.256 MAL13P1.256 phosphatidylinositol transfer protein, putative 19107032:19112954 forward M	MAL13P1.256	CM, ABM, EN	CM, ABM		0005622 : intracellular	0006810 : transport	*Plasmodium falciparum*
MAL7P1.167 MAL7P1.167 hypothetical protein, conserved 7977659:7985980 reverse MW:329471	MAL7P1.167	CM, ABM, EN	CM, ABM				*Plasmodium falciparum*
PF10_0267 PF10_0267 hypothetical protein 12217392:12218954 forward MW:61468	PF10_0267	CM, ABM, EN	CM, ABM				*Plasmodium falciparum*
PF11_0374 PF11_0374 hypothetical protein 14189790:14193247 forward MW:128546	PF11_0374	CM, ABM, EN	CM, ABM	0003676 : nucleic acid binding			*Plasmodium falciparum*
PF14_0364 PF14_0364 cleavage and polyadenylation specifity factor protein, putative 21529938:215325	PF14_0364	CM	CM, ABM	0016787 : hydrolase activity			*Plasmodium falciparum*
PFD0535w PFD0535w hypothetical protein 3143210:3147106 forward MW:150271	PFD0535w	CM, ABM	CM				*Plasmodium falciparum*

5.3.2 P rotein Characterisation and Categorisation

Tables found in **Appendices VII-X** give a summary of all the proteins that had corresponding GO numbers. 159 *human* and 288 *P. falciparum* proteins identified from the database searches of the plasma samples were catalogued. 102 *human* and 62 *P. falciparum* proteins identified from the CSF spectra were catalogued using the GO annotations. **Figures 5.1-5.3** give a graphic distribution of the 26 *P. falciparum* proteins unique to plasma drawn from patients diagnosed with CM. 2 of the 3 human proteins found unique in the plasma (**Table 5.2**) had corresponding GO functions and both were associated with protein binding. **Figures 5.4-5.6** give a graphic distribution of the 8 *human* proteins unique to the CSF drawn from children diagnosed with CM (**Table 5.3**). The 2 *P. falciparum* proteins were hypothetical and were not analysed. 8 *human* and 6 *P. falciparum* (**Table 5.4**) proteins were found in both plasma and CSF drawn from children diagnosed with CM and graphic distributions using GO categories are shown on **figures 5.7-5.12.**

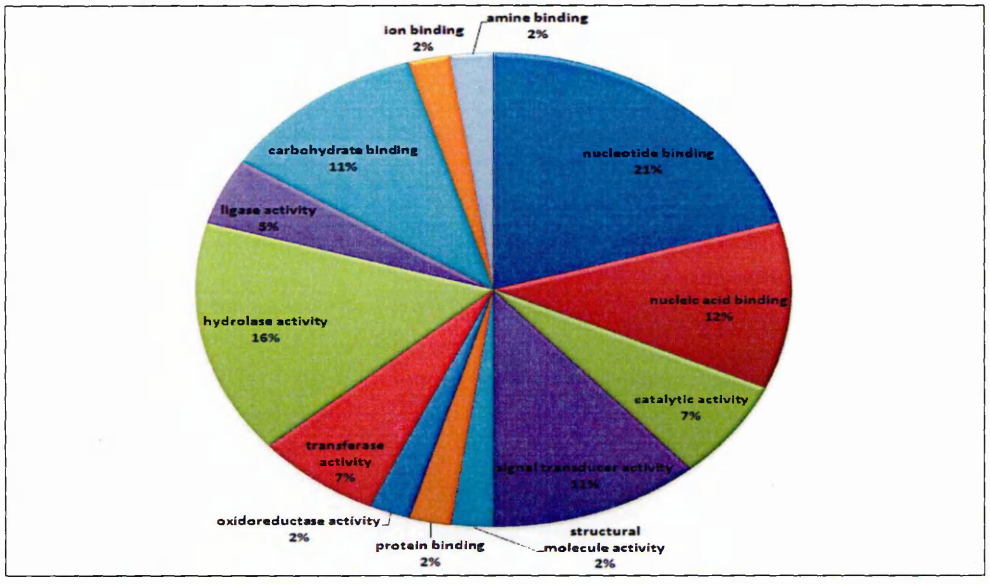

Figure 5.1 Distribution of 26 *P. falciparum* proteins found only in plasma drawn from children with CM, **categorised according to gene ontology (GO) annotation for molecular function.**

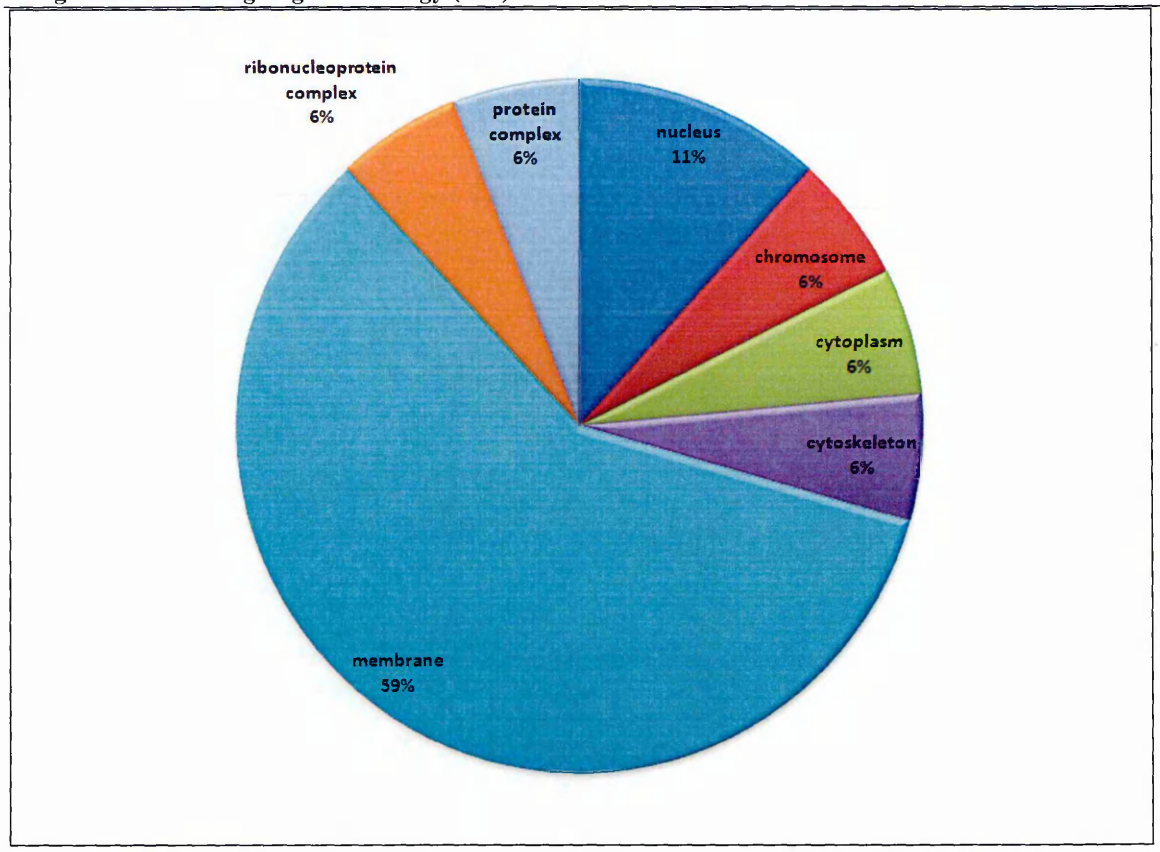

Figure 5.2 Distribution of 26 *P. falciparum* proteins found only in plasma drawn from children diagnosed CM. The proteins were categorised according to gene ontology (GO) annotation for cellular component.

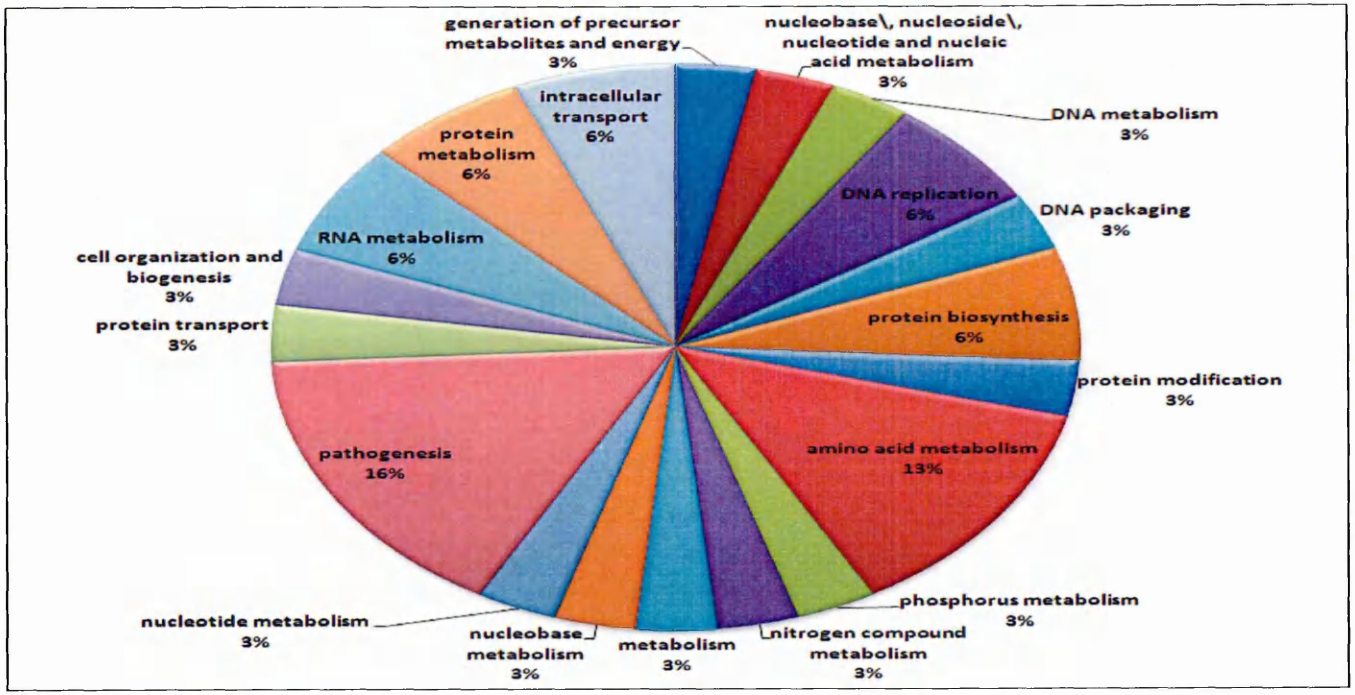

Figure 5.3 Distribution of 26 *P. falciparum* proteins found only in plasma drawn from children diagnosed with CM. The proteins were categorised according to gene ontology (GO) annotation for biological process.

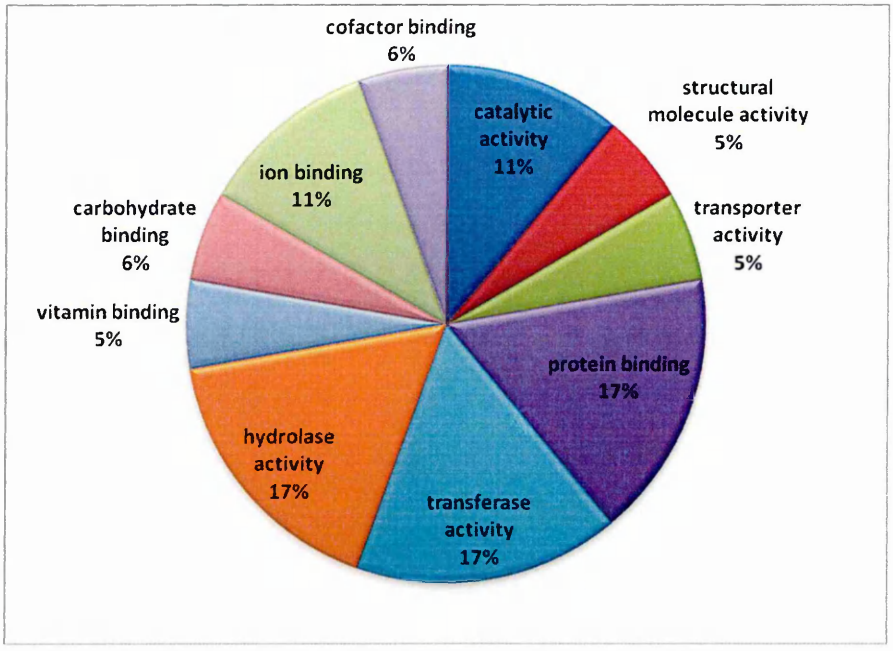

Figure 5.4 Distribution of 8 *human* proteins found only in CSF drawn from children with CM, categorised according to gene ontology (GO) annotation for molecular function.

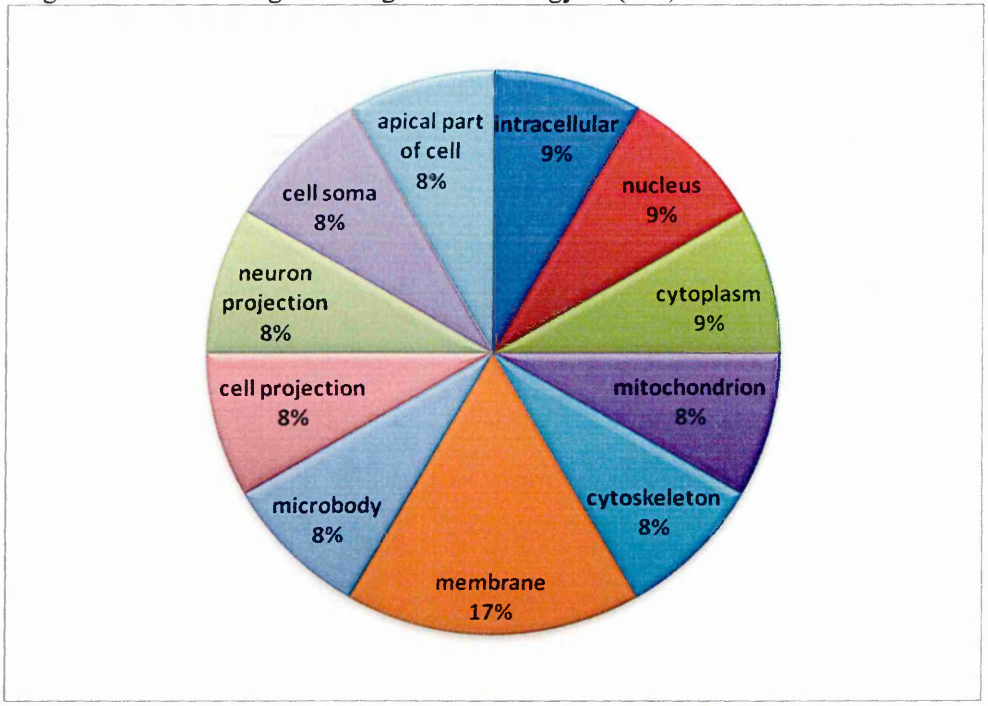

Figure 5.5 Distribution of 8 *human* proteins found only in CSF drawn from children with CM categorised according to gene ontology (GO) annotation for cellular component.

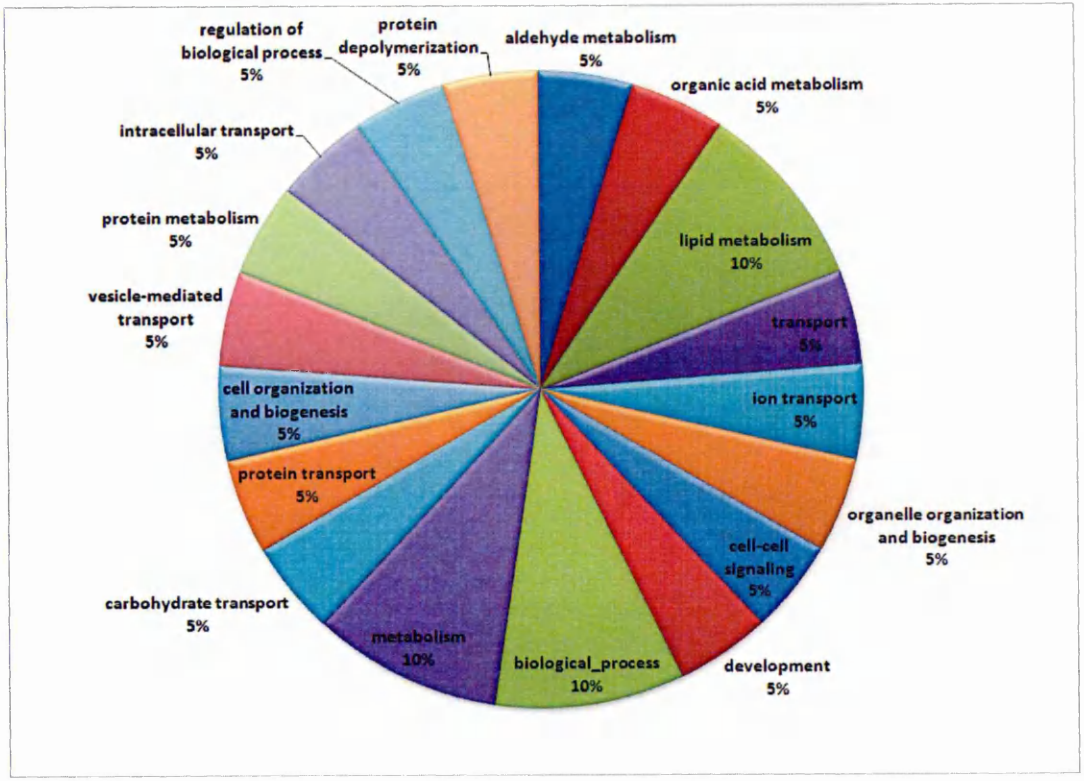

Figure 5.6 Distribution of 8 *human* proteins found only in CSF drawn from children diagnosed with CM. The proteins were categorised according to gene ontology (GO) annotation for biological processes

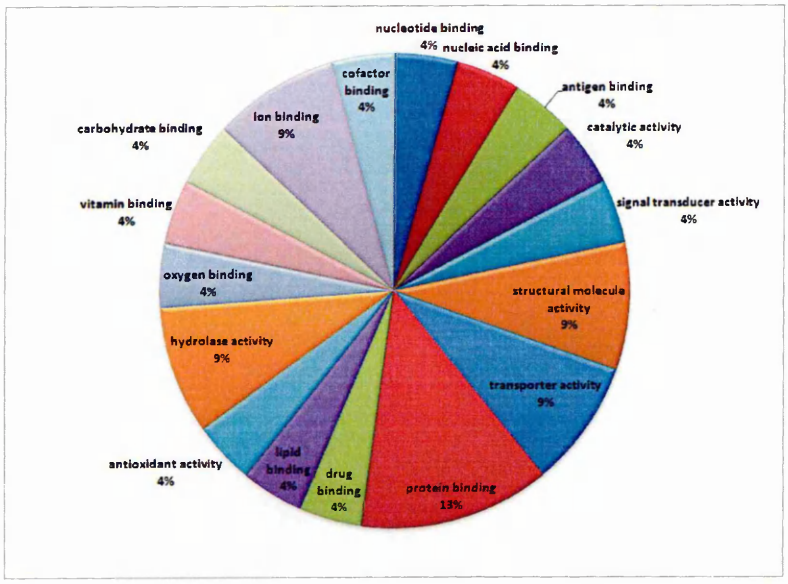

Figure 5.7 Distribution of 8 *human* proteins found in both CSF and plasma drawn from children with CM categorised according to gene ontology (GO annotation for molecular function.

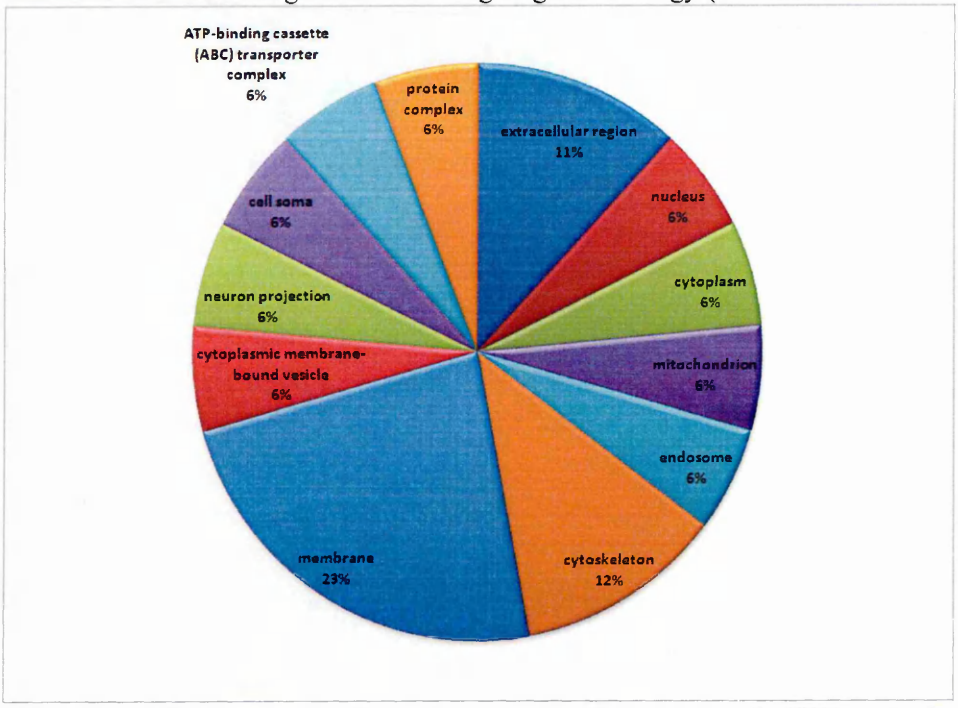

Figure 5.8 Distribution of 8 *human* proteins found in both CSF and plasma drawn from children with CM categorised according to gene ontology (GO annotation cellular component.

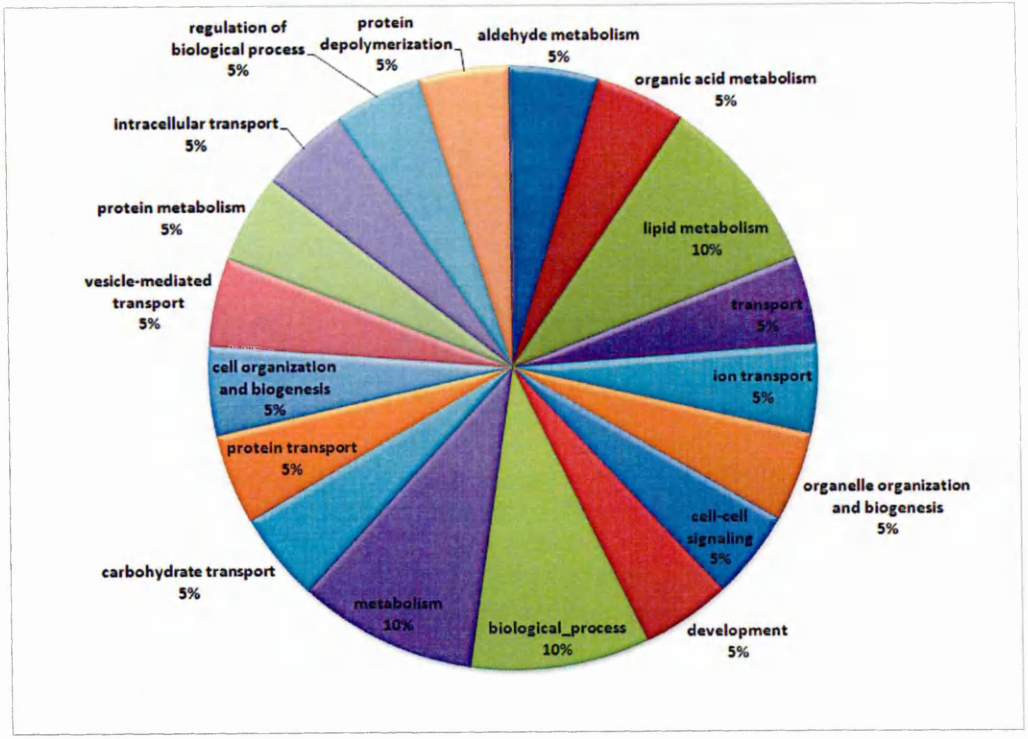

Figure 5.9 Distribution of 8 *human* proteins found in both CSF and plasma drawn from children diagnosed with CM. The proteins were categorised according to gene ontology (GO) annotation for biological process.

5.4 Discussion

Protein expression profiles in CSF and plasma have previously been produced through a combination of 1D and 2D gel electrophoresis (2-DE) and MS. Despite being a powerful tool of protein separation 2-DE combined with MS has a limited dynamic range and is not sensitive enough for identification of the LAPs in complex samples. Another major disadvantage is the time and labour needed for spot-by-spot analysis making it less suitable for rapid large scale analysis of complex protein mixtures. A substitute for 2-DE should therefore allow for the rapid identification of proteins and deal equally well with all proteins regardless of their abundance in the sample.

Non gel based approaches such as MudPIT have become a real alternative for high through put proteomic investigations. A higher number of proteins are identified in non-gel based methods a good example being yeast samples where 2-DE has been reported to resolve ~1200 proteins (Futcher et al., 1999) but only 148 (Futcher et al., 1999) and 279 (Perrot et al., 1999) of the resolved proteins have been identified by MS. By contrast, the MudPIT technology has resolved and identified 1484 proteins (Washburn et al., 2001). However, the abundances of individual proteins may range over several orders of magnitude in a crude mixture thus providing an analytical challenge especially when MudPIT is used. Consequently, high abundance proteins such as albumin interfere with the detection of other proteins degrading the overall performance of the process (Ahmed et al., 2003, Butt et al., 2001).

Complexity of starting sample plays a key role in protein profiling of biological samples. Pre-fractionation of the sample (by SDS-PAGE or liquid chromatography steps or sub cellular fractionation) can help reduce the sample complexity and this improves the

resolving power of the analysis. A variety of multi dimensional chromatographic steps have been described to replace 2-DE based protein separations prior to MS (Liu et al., 2002). In this chapter the use of LC-MS/MS as an alternative to 2-DE and a comparison of various separation techniques were explored. The one that gave the most data in the least amount of time was then chosen for subsequent work. The MudPIT approach (Link et al., 1999, Mitulovic et al., 2004) was initially tried, and resulted in only 35 proteins being identified in a CSF sample. With the modified shotgun method the samples were first fractionated prior to enzyme digestion resulting in a higher number of proteins. The significant difference can be explained by the fact that trypsin digestion is more efficient when smaller numbers of proteins are presented which influences the subsequent peptide fragmentation data retrieved. Another major difference in the method described in this chapter is that instead of using a SCX column for the pre-fractionation step it was found that using the monolithic column resolved albumin and other abundant proteins better and albumin was only found in 2 fractions compared to 4 or more fractions when SCX was used as a first dimension column. Furthermore, the albumin isoforms were all in the same fraction and not separated as with the SCX. It is worth mentioning that the number of proteins identified increased with the number of fractionation steps done prior to LC-MS/MS but this would increase the time taken to analyse one sample reducing the high through put capability of the technique. It should also be taken into account that the more dimensions performed the higher the amount of loading sample one would need to ensure that there is enough material to characterise low abundant proteins.

For the second dimension separation, a monolithic silica based column rather than a RP column was used. The former had the added advantage of shorter gradients of 30 minutes

compared to 60 minutes on the RP column. Use of these columns has also been noted to enhance chromatographic resolution (Tanaka et al., 2001, Motokawa et al., 2002, Ishizuka et al., 2002, Walcher et al., 2002), providing a perfect means by which to reduce ion suppression.

Several CSF studies described in literature with used pooled patient samples. However this approach does not give a clear picture of what protein profiles are constant in each disease manifestation. In the study described in this chapter, samples were not pooled, which greatly limited the amount of starting material but ensured that any protein profiles identified were actually present in all patient samples.

Bioinformatics plays an essential role in a successful proteomic analysis. Access to the fully sequenced genomes of the *human* and *P. falciparum* database has enabled the identification of many proteins from our samples. The TurboSequest® algorithm based on the SEQUEST® algorithm described by Yates (Yates, 1998, Link et al., 1999, Yates et al., 1999, Washburn et al., 2001) was used in this study and it enabled us to search all our data against databases of the *human* and *P. falciparum* subsets of NCBI and the *P. falciparum* database downloaded from the Sanger Institute. However, when data from the two *P. falciparum* databases were compared, there were some differences, which concurs with findings from other studies (Pan et al., 2007, Carr et al., 2004) and therefore any study should standardise the database used to avoid differences.

Despite protein amount normalisation (200ug in plasma and 100ug in CSF), a significant number of non-hypothetical human proteins in plasma (76%) and CSF (66%) were found only in the ABM diseases group. This would suggest that there are quite a few proteins

that are either present in lower quantities or completely absent in the other 2 disease groups and quantitative analysis would have to be done to determine this. Orosomucoid alpha glycoprotein, one of 2 proteins found unique to the plasma from the CM group had also been found unique to the CM plasma gels in the previous chapter. This protein is an acute phase protein and has previously been studied in connection with malaria (Beesley et al., 2000). However, one would expect to see it in ABM too as it is associated with inflammation. Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3) was identified as unique to CSF samples drawn from patients diagnosed with CM. This protein was also found to be uniquely present in infected mouse plasma in chapter 2. One study has recently found a correlation between CM and the presence of non-erythoid alpha spectrin protein in Gabonese children (Guiyedi et al., 2007). However the study also found that the alpha chain was more abundant than the beta chain in the brain material studied. The repertoire of brain antigens recognized by plasma IgGs was more diverse in infected than in uninfected individuals. Anti-brain reactivity was significantly higher in the CM group than in the uncomplicated malaria and severe non-CM groups. IgG self-reactivity to brain antigens was also correlated with plasma IgG levels and age. 90% of CM patients displayed reactivity to a high-molecular mass band containing the spectrin non-erythroid alpha chain. Reactivity with this band was correlated with high TNFα concentrations in CM patients. These results strongly suggest that an antibody response to brain antigens induced by *P. falciparum* infection may be associated with pathogenic mechanisms in patients developing CM.

Out of the 323 *P. falciparum* proteins identified in the plasma, 15 were common to all three disease groups and these could represent proteins normally found in this malaria population. However, a very high number of *P. falciparum* proteins were found in both CM and ABM (66) and another 154 unique to ABM. However, this is not unexpected as all children in the study area are exposed to malaria, with up to 300 infected bites per year. The high number of *P. falciparum* proteins in ABM also correlates to a high number of *human* proteins identified in this disease group and this would suggest that although there was normalisation for protein amount, the proteins are either present in too low a concentration to determine with this method or a larger number of post translational modification of proteins is occurring in the CM and EN group hence making it harder to get protein identifications with this technique. Further analysis using other techniques would be needed to understand these results.

5.5 Conclusion

The ability to identify a significant number of proteins related to severe disease (typically low in abundance) is largely due to the utilisation of proteomic platforms based on different separation strategies as well as mass spectrometers with different ionisation sources and mass analyzers. The benefits of the multi-dimensional orthogonal sample separation described in this chapter far outweigh the use of 2-DE. However, to completely eliminate the use of gel based methods there is a need to incorporate quantification of the proteins in the LC system. One way of doing this is by use of reagents such as ICAT which tags proteins containing cysteine residues only, which also enhances the analytical dynamic range, and thus, improves the identification of low-abundance proteins (Zhang et al., 2005a). Results in this chapter have demonstrated that 'shotgun' sequencing

approaches (LC-MS/MS) not only can profile high-abundance proteins in complex biological fluids but also have the potential to identify some low-level proteins present in such complex mixtures without extensive pre-purification protocols. A key to such studies, however, is to use targeted approaches that reduce the complexity of the solute mixture that is presented to the mass spectrometer at a given time point.

It would be interesting to determine the prevalence and levels in endemic residents of anti-auto-antibodies of the 2 proteins found unique in plasma of CM patients. If two cohorts distinguished by significantly different levels of these antibodies could be constituted, then it will be possible to conduct a prospective study that could allow determining whether high autoantibody levels predispose an individual to develop CM.

6 Mass Profile of Proteins Associated with Cerebral Malaria in Children

6.1 Introduction

Molecular profiling experiments performed directly on tissues using MALDI mass spectrometry are gaining considerable popularity (Rohner et al., 2005, Reyzer and Caprioli, 2005, Crossman et al., 2006, Altelaar et al., 2006, Garrett and Yost, 2006). When proteins are added to a matrix solution such as α-cyano-4-hydroxycinnamic acid, they become charged and can be detected as peaks in a mass spectrum with unique mass to charge ratio (*m/z*) values. Since most of the proteins are singly charged during the analysis, each peak usually corresponds to a single protein with a molecular weight equivalent to the *m/z* value. The normalised peak intensity is directly proportional to the concentration of the corresponding protein in the sample. MALDI-ToF MS offers a technique for rapid determination of molecular masses and the heterogeneity of small amounts of peptides and proteins (Karas et al., 2000, Sottani et al., 1997, Belgacem et al., 2002, Sekiya et al., 2005). The technique is characterised by a number of useful features including minimal sample preparation, ease of use, heightened throughput, and cellular specificity. It is increasingly being used to refine disease characterization in some patient populations. Recently, for example, there has been a great deal of interest in using this technique to try and identify quantitative or qualitative differences in serum protein components between cancer patients and control subjects. The existence of such differences is postulated on the basis that, when cancer cell proliferation products such as tumor-specific proteins enter the circulation, they change the profile of circulating serum and/or plasma proteins. Serum profiling (i.e., the characterisation of proteins, peptides,

and macromolecules from serum) by surface enhanced laser desorption/ionization (SELDI) mass spectroscopy (Petricoin and Liotta, 2004) coupled with statistical algorithms has been used to distinguish patients with cancer from control subjects and patients with benign conditions (Adam et al., 2002, Petricoin et al., 2002c, Valerio et al., 2001). However, the use of SELDI involves long and expensive preparative steps with potential losses of selected protein components (Hortin and Remaley, 2006)

In contrast, application of MALDI-ToF MS in medical research has great promise because i) fingerprints will help define *disease states* rather than just pathogen detection, ii) the analysis requires no prior assumptions about the nature of proteins, iii) protein identification is not essential for diagnostic utility and iv) it accommodates biological heterogeneity in disease expression.

The primary aims of this study were to evaluate methods of CSF and plasma processing for MALDI-ToF analysis and to use MALDI-ToF combined with bioinformatics analysis in a proof of principle study to describe a mass profile of proteins in plasma and CSF from children with CM. These profiles were compared to mass profiles of samples collected from children with EN, ABM, severe malaria with seizures (SZ) and prostate children with severe malaria (PR)

6.2 Methodology

The general workflow for this study is summarised in **Fig. 6.1** and the background to these methodologies is described in detail in chapter 2 section 2.4. In brief the technique used involves (1) sample preparation (2) MALDI-ToF in linear mode (3) qualitative analysis of peaks of interest.

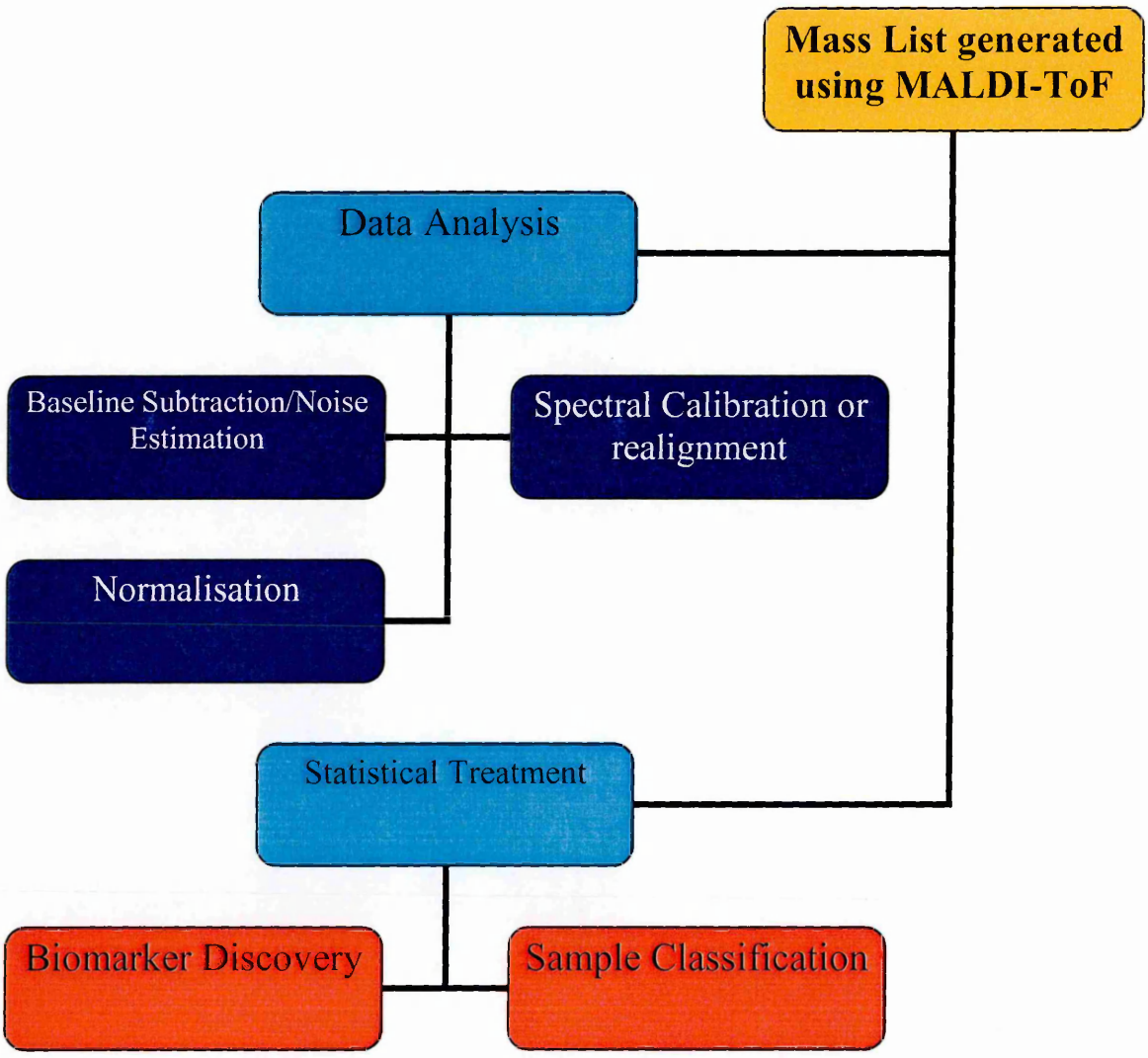

Figure 6.1 Mass Spectral Analysis Work Flow. The mass spectra generated using MALDI-ToF are treated to processing algorithms responsible for the removal of noise, realignment of the *m/z* scale, and peak selection and matching. The mass list generated is exported to a table and formatted for qualitative analysis using a number of established methods. The result of the analysis is a list of biomarker candidates that are subjected to further validation steps.

6.2.1 Sample Source

Plasma and CSF samples collected from children in Kilifi with different presentation of malaria–Cerebral Malaria (CM), Malaria+ seizures (SZ), Malaria + prostate (PR)were compared to samples collected from children having a final diagnosis of ABM and EN (n=12 for each group). Clinical details are given shown on **Table 6.1**

Table 6.1 Clinical details of patients used in mass profiling of proteins

	Cerebral Malaria (CM) (n=12)	Acute Bacterial Meningitis (ABM) (n=12)	Undiagnosed Encephalopathy (EN) (n=12)	Severe Malaria with Seizures (SZ) (n=12)	Prostate Children with Severe Malaria (PR) (n=12)
Age[10]	*33.89 (21.75)*	*52.41 (51.33)*	*24.85 (15.64)*	*32.81 (38.46)*	*31.69 (17.32)*
Sex (%Male)	*33*	*42*	*33*	*42*	*25*
Duration of illness before admission to hospital[11]	*2.5 (1.31)*	*2 (1.18)*	*2*	*2.63 (1.61)*	*6.25 (9.6)*
Number with Seizures	*8*	*8*	*10*	*12*	*0*
Number with Coma	*11*	*5*	*8*	*2*	*4*
Hb (s.d.)	*6.89 (1.94)*	*8.6 (2.4)*	*10 (1.05)*	*8.3 (1.45)*	*7.2 (3.12)*
Duration in Hospital[12]	*4.4 (2.6)*	*34 (54.50)*	*4.7 (3.04)*	*4.09 (3.81)*	*4.12 (1.8)*
Mortality (%)	*0*	*33*	*17*	*0*	*0.08*

[10] Average age in months (s.d)
[11] Average number of days (s.d)
[12]Average number of days (s.d.)

6.2.2 Sample Preparation

Protein determination was determined as described in chapter 2 section 2.1.2.1. Samples were then prepared for mass spectrometry as described in chapter 2 section 2.4.2.1

6.2.3 Mass Spectrometry

Mass spectra were generated using an AXIMA CFR Plus™ (Shimadzu Biotech-Kratos Analytical, Manchester, UK) mass spectrometer operated in a linear mode_2ghz which was set as described in chapter 2 section 2.4.2.2. After baseline subtraction and normalization using the Kratos software provided with the MALDI-ToF MS, a mass list of all masses present from 500-180,000 Daltons (Da) was generated for each sample.

6.2.4 Qualitative Analyses

The mass list generated from the MALDI-ToF was then exported on to an excel sheet. The mass and intensity of the peaks were then transported to STATA and the masses sorted and masses below 2000kD dropped from the plasma samples. Samples from each patient in a disease phenotype were compiled into one file and representative spectra were then generated using R© version 2.5.1 (2007-06-27) (http://www.r-project.org/) and the scripts for generating the spectra are attached as Appendix XI.

6.3 6.3 Results

Figures 6.2-6.11 show representative spectra per disease group. Despite sample dilution and protein normalisation prior to spotting of the MALDI target, the spectra yielded from the plasma had very few peaks of high molecular weight proteins.

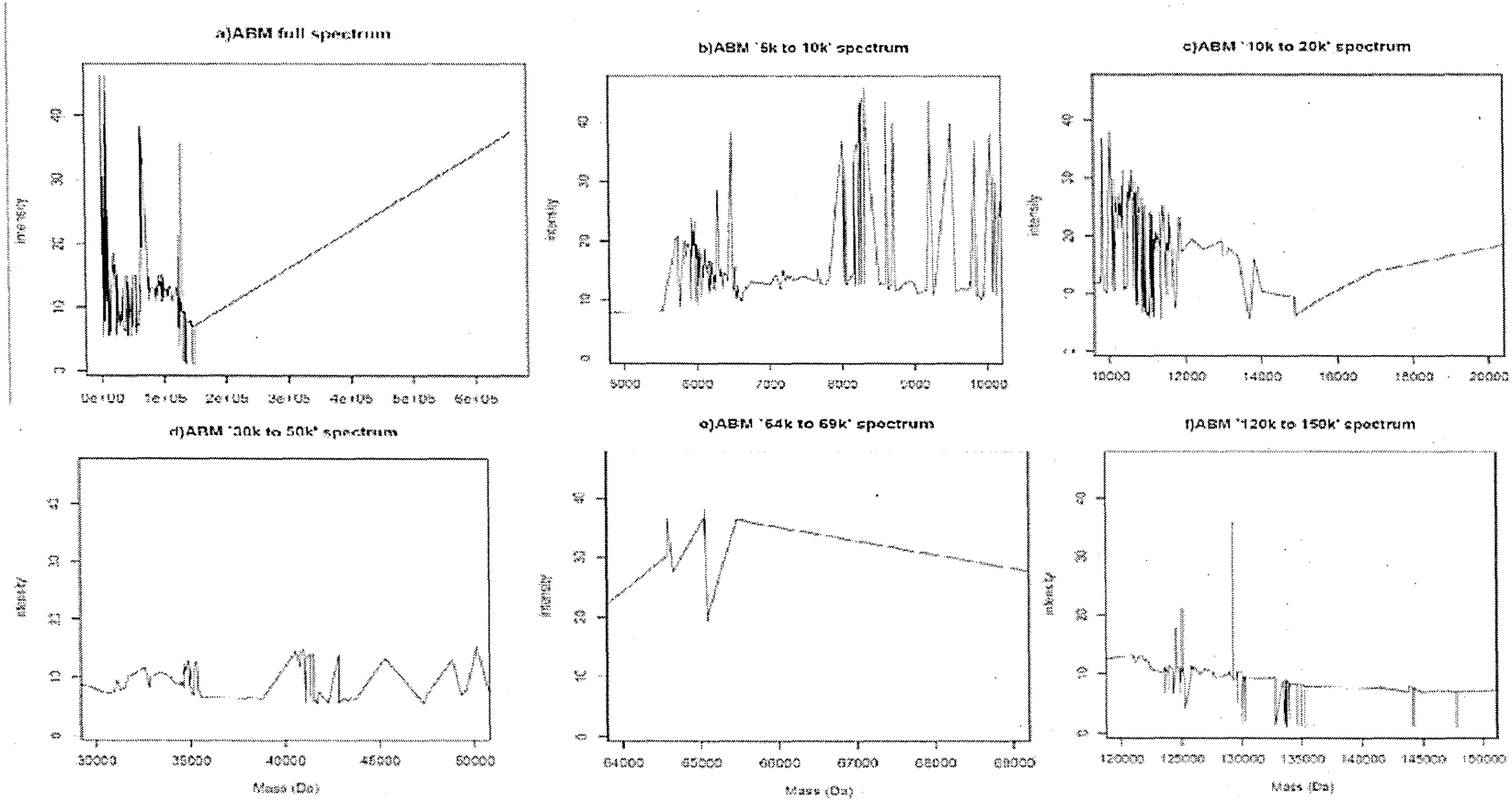

Figure 6.2 Representative spectra for plasma drawn from children with ABM. Spectrum a) shows the full mass range and b-f show areas of interest.

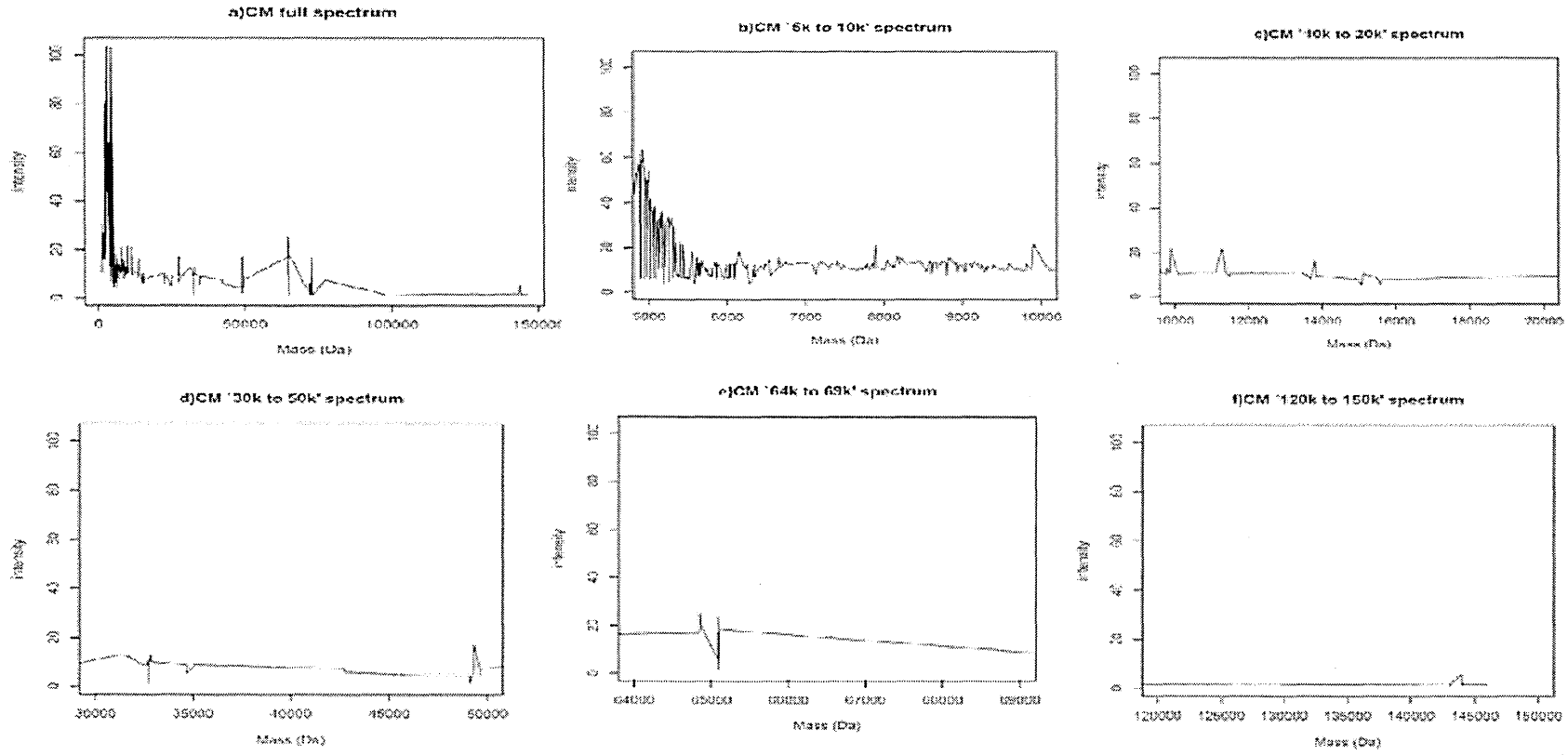

Figure 6.3 Representative spectra for plasma drawn from children with CM. Spectrum a) shows the full mass range and b-f show areas of interest.

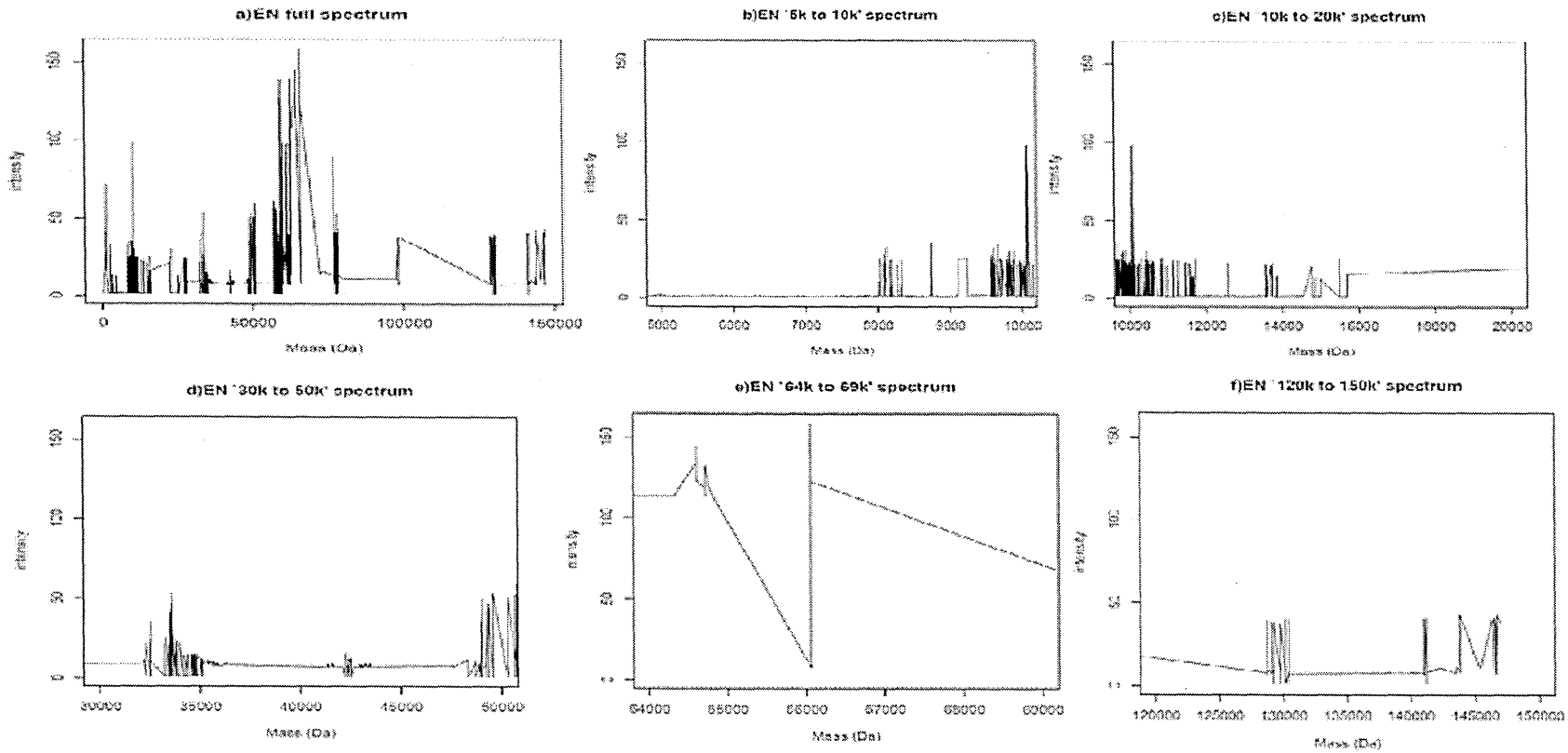

Figure 6.4 Representative spectra for plasma drawn from children with EN. Spectrum a) shows the full mass range and b-f show areas of interest.

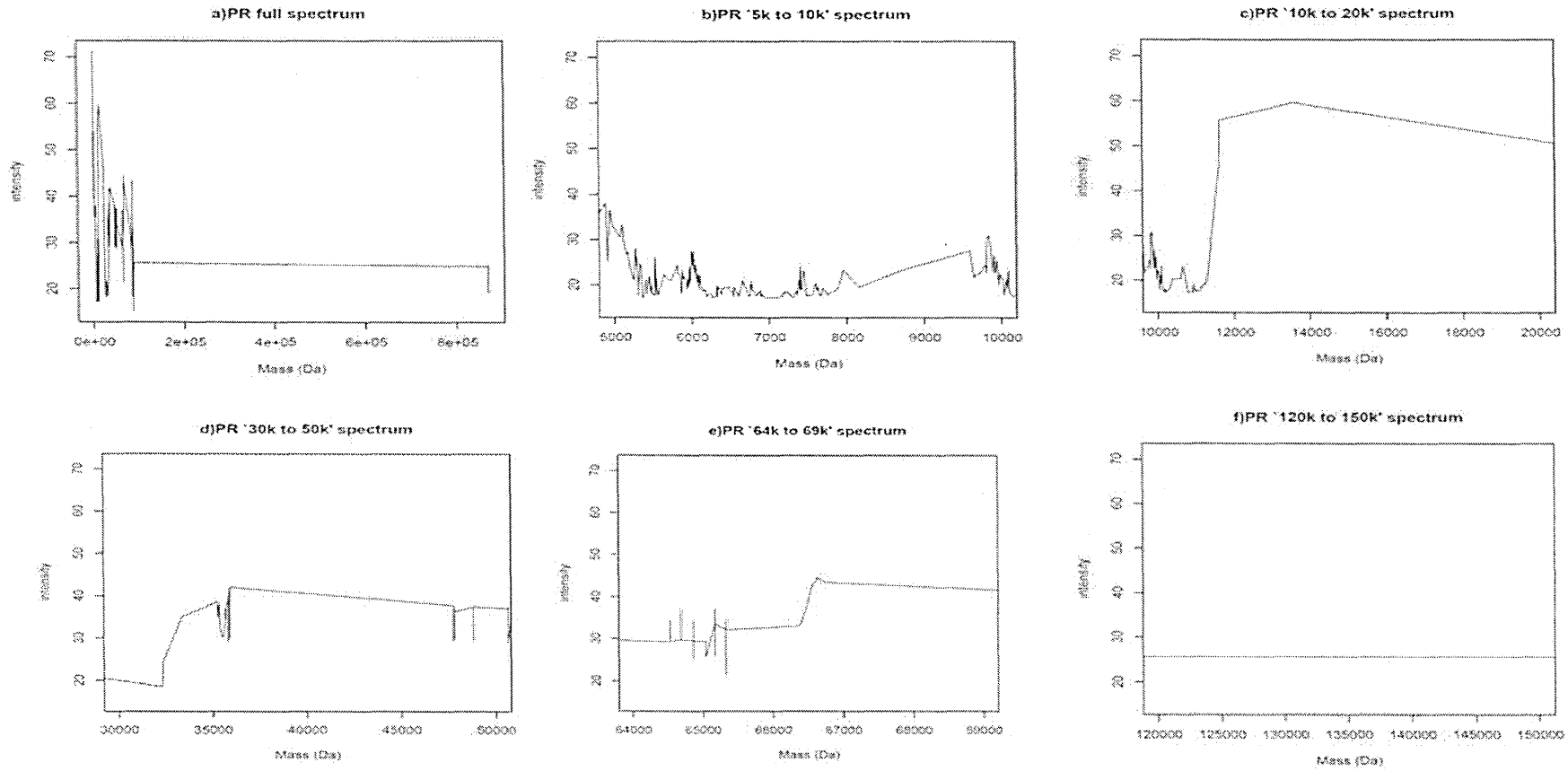

Figure 6.5 Representative spectra for plasma drawn from children with PR. Spectrum a) shows the full mass range and b-f show areas of interest.

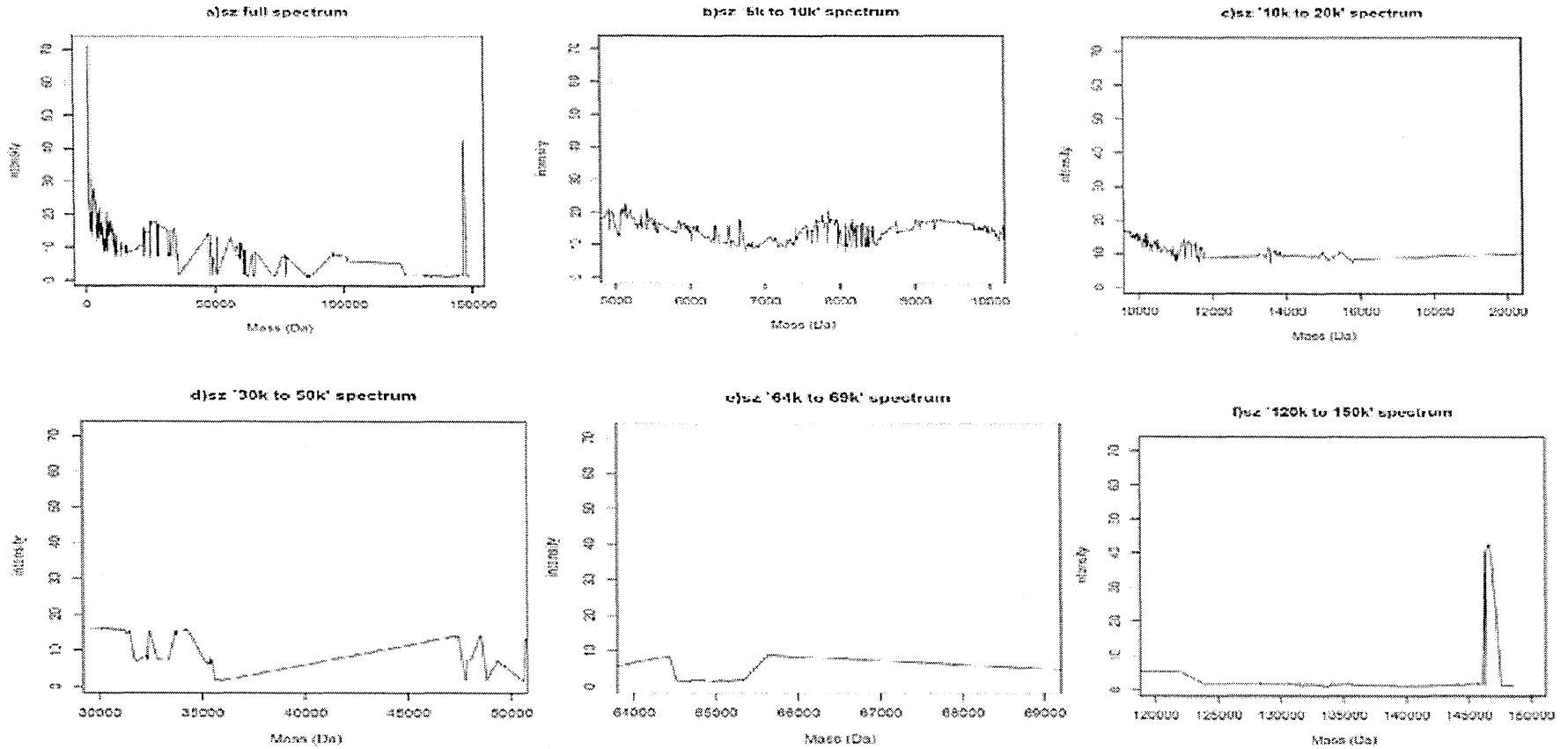

Figure 6.6 Representative spectra for plasma drawn from children with SZ. Spectrum a) shows the full mass range and b-f show areas of interest.

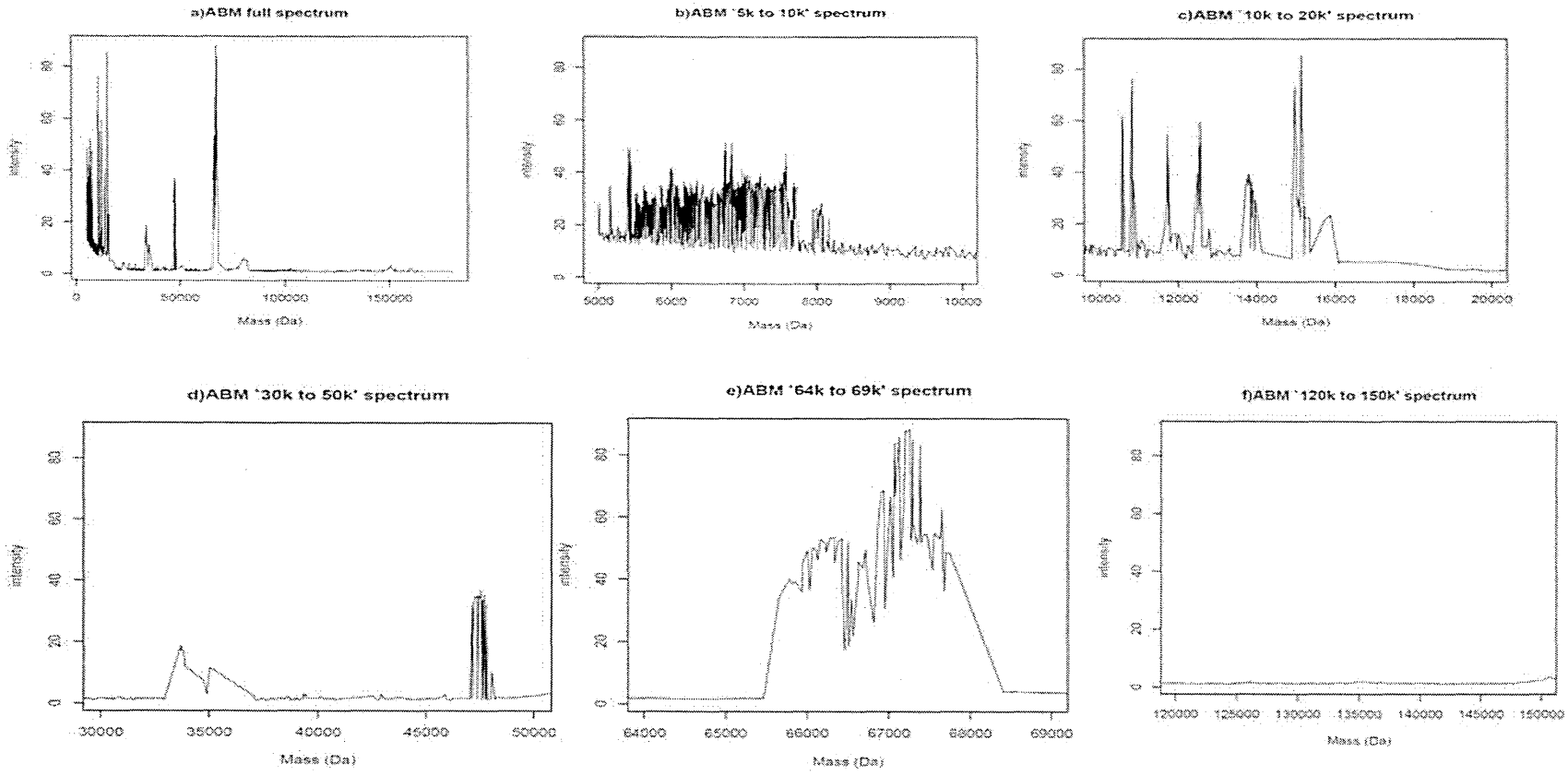

Figure 6.7 Representative spectra for CSF drawn from children with ABM. Spectrum a) shows the full mass range and b-f show areas of interest.

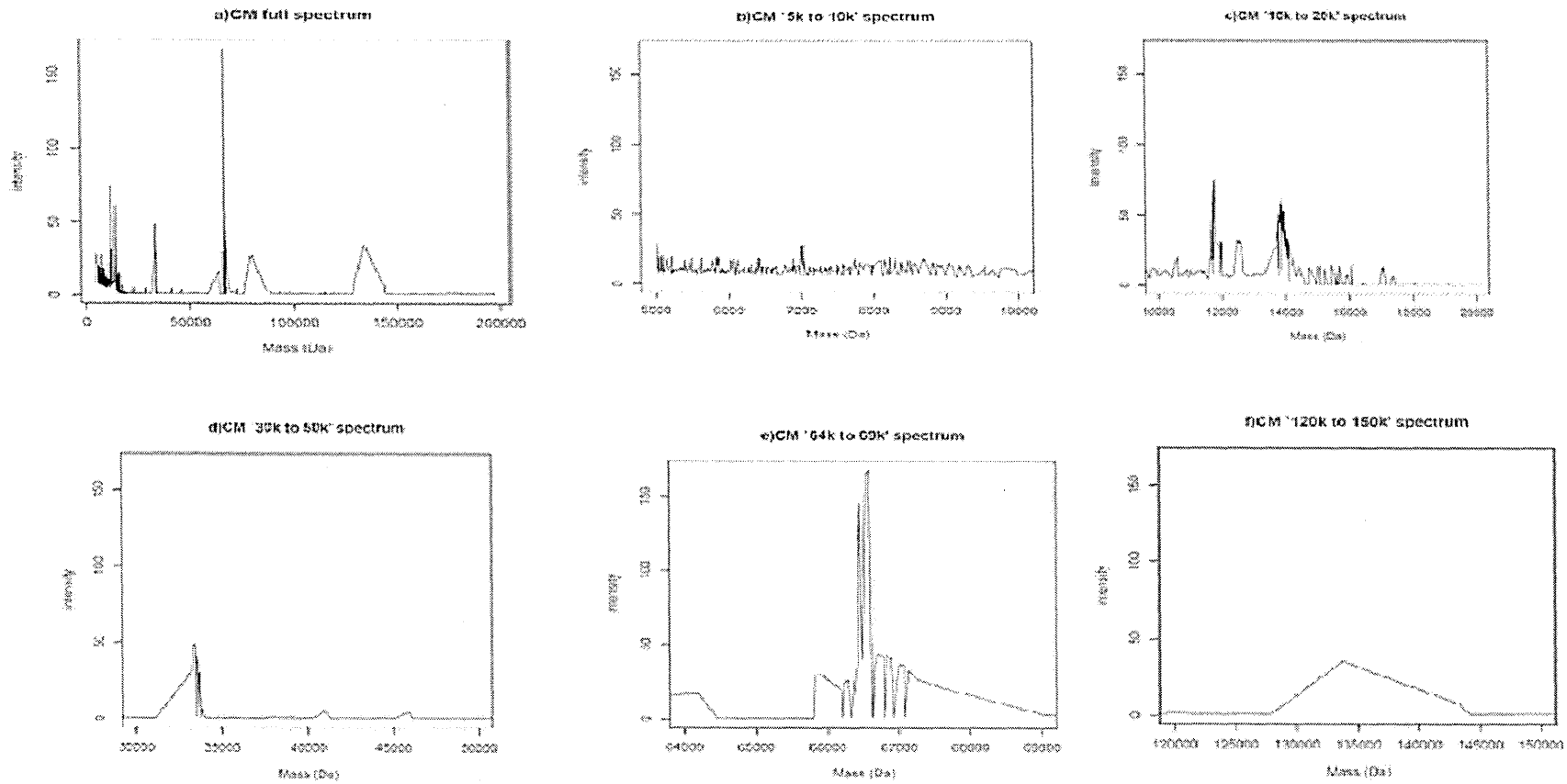

Figure 6.8 Representative spectra for CSF drawn from children with CM. Spectrum a) shows the full mass range and b-f show areas of interest.

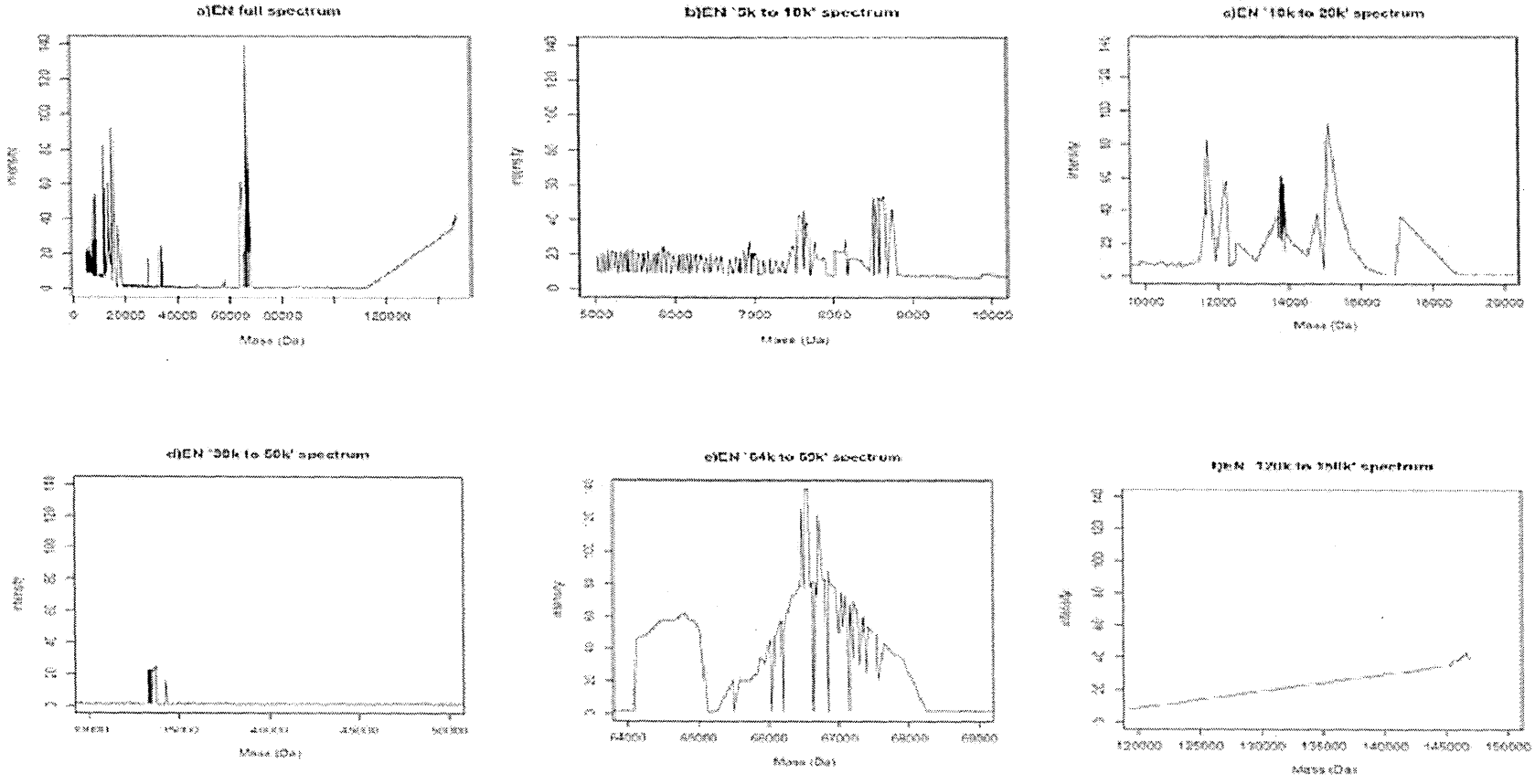

Figure 6.9 Representative spectra for CSF drawn from children with EN. Spectrum a) shows the full mass range and b-f show areas of interest.

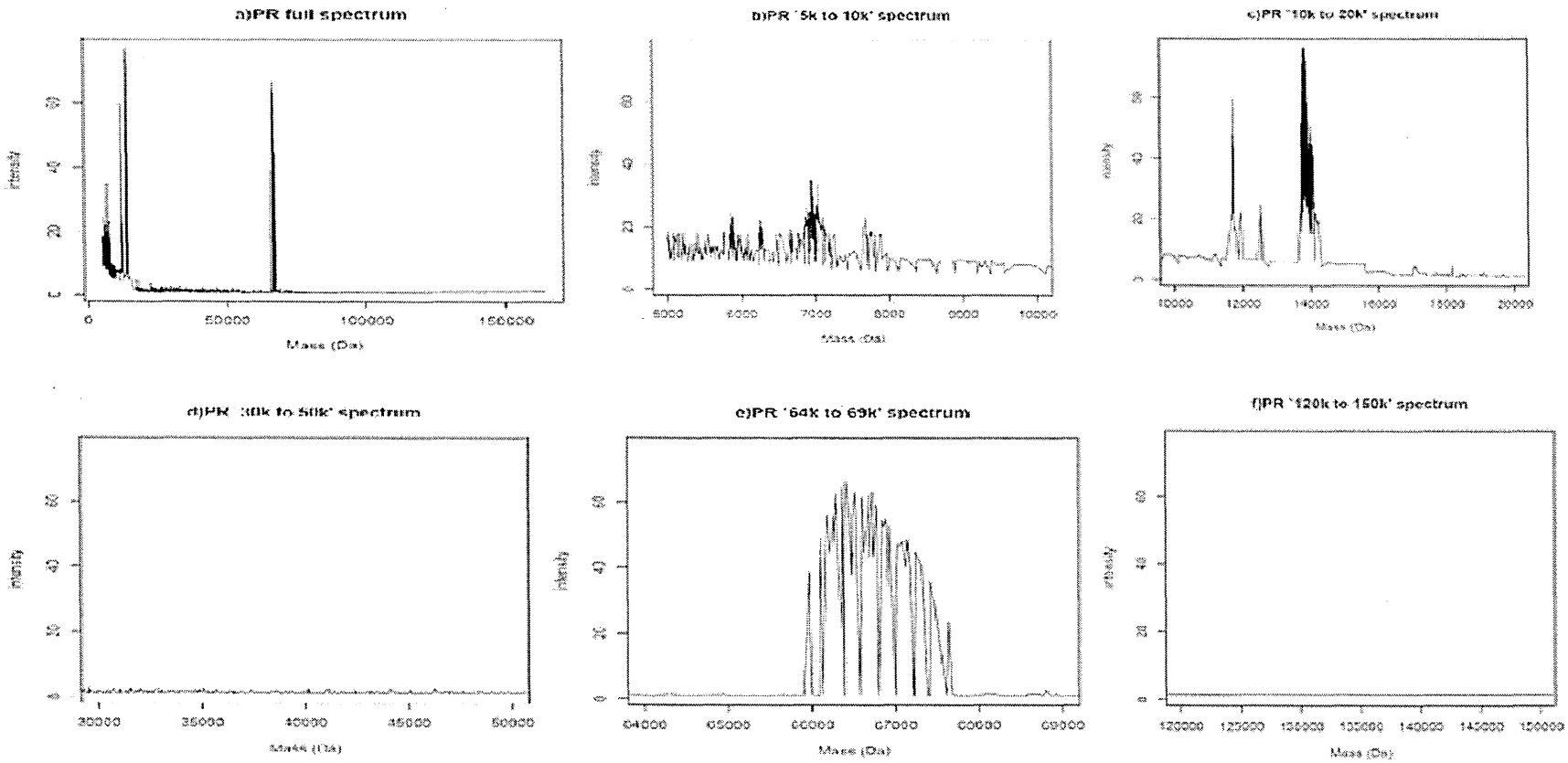

Figure 6.10 Representative spectra for CSF drawn from children with PR. Spectrum a) shows the full mass range and b-f show areas of interest.

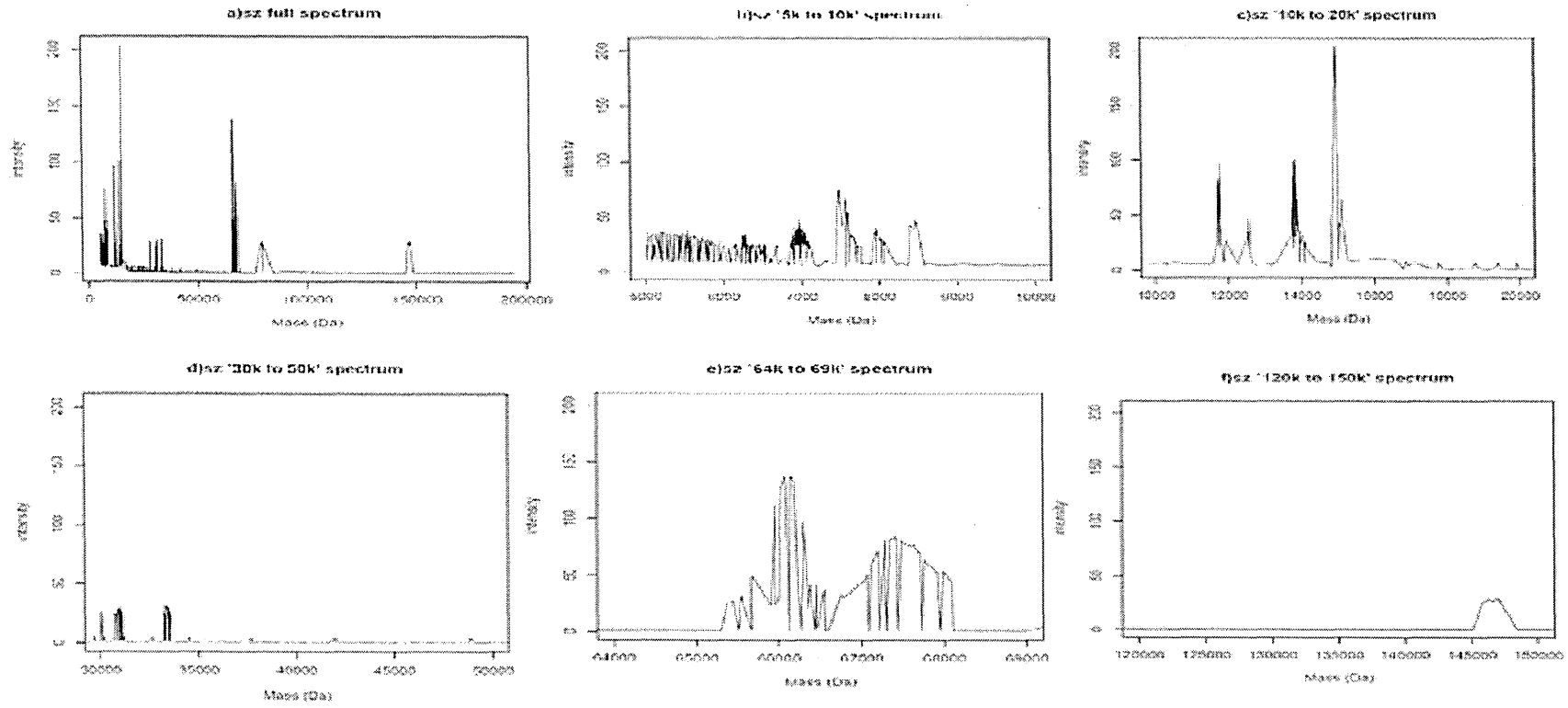

Figure 6.11 Representative spectra for CSF drawn from children with PR. Spectrum a) shows the full mass range and b-f show areas of interest.

6.4 Discussion

Disease marker discovery is one of the most promising applications of MS in biomedical research (Petricoin et al., 2002c, Steel et al., 2003, Rai et al., 2002, Li et al., 2002b, Adam et al., 2002). In addition to structural analysis of proteins , MS is now being used in biomedical research to search for biomarkers in extremely complex mixtures (Petricoin et al., 2002c, Steel et al., 2003, Rai et al., 2002, Li et al., 2002b, Adam et al., 2002), quantify changes in protein abundance (Gygi et al., 1999, Sinz et al., 2002), and generate profiles for diagnosis of different disease states by direct analysis of proteins extracted from biological fluids (Sinz et al., 2002, Catinella et al., 1999), cells (Palmer-Toy et al., 2000, Xu et al., 2002), and tissues (Caprioli et al., 1997, Stoeckli et al., 1999). SELDI, the coupling of surface retentate chromatography with MALDI MS, is already widely used for clinical research because of the convenient packaging of a low resolution, automated ToF mass spectrometer with ready-made ''chips'' modified with different functional groups for performing surface retentate chromatography (Weinberger et al., 2002, Issaq et al., 2003). However, findings from re-analysis of the raw data of some of the studies suggested that the identified diagnostic SELDI peaks could be due to artifact of sample processing (Baggerly et al., 2004, Baggerly et al., 2005). Moreover, where the protein peaks have been identified, they have turned out to be well-known, acute phase proteins. Critics of SELDI have claimed that it is inherently limited in its depth of coverage, with a dynamic range that prevents the detection of low abundant proteins (Diamandis, 2004a, Diamandis, 2004c). As for predictive patterns, no two studies have found the same patterns to date. In addition, none of the patterns have been validated in an independent study by another laboratory. In one study, both the sensitivity and the specificity of the pattern declined

significantly when samples were processed in the same laboratory after a delay of several months (Rogers et al., 2003). Analysis of serum or plasma by SELDI-TOF MS offers some fractionation of components before MS, but unless immunocapture is applied, the efficiency of fractionation generally is modest, so that the major serum or plasma components predominate (Rossi et al., 2006).

In the study in this chapter, unfractionated plasma and CSF have been analyzed by MALDI-ToF MS. The observed profiles of peaks for unfractionated CSF specimens are largely in agreement with reported profiles of plasma proteins. Despite dilution, the plasma proteins did not ionise as well as expected probably due to over abundance in albumin leading to low matrix to sample ratio. Another problem noted was that despite baseline subtraction there was still a major drift in the spectra from plasma which made data analysis difficult. Representative spectra from the CSF were generally better consistently yielding a few peaks in the high mass range as well as yielding sharper less broad peaks suggesting that the ratio of high abundant proteins to matrix provided a more efficient ionisation of the sample. Another factor that may have led to better spectra of the CSF was the fact that the CSF samples had been desalted prior to analysis. The CSF spectra yielded similar patterns but a few differences were noted. In the ABM spectra there were many peaks between 5 kD -8kD compared to the other disease phenotypes. All phenotypes had a peak at around 34kD-36kD except PR. In literature a mass varying between 32kD and 36 kD is normally attributed to α_1-acid glycoprotein (Chait and Kent, 1992, Sekiya et al., 2005). The range in masses is explained by the fact that loss of sialic acid may vary depending on technique of analysis. All spectra had abundant masses ranging from 64kD to 69kD which are attributed to albumin. However the peaks in the

CSF from the CM group were not as an abundant as in the other group and only 2 well defined peaks at around 66500 were seen. The SZ group had a peak at 146kD not seen in other disease phenotype.

The CSF profiles are dominated by the most abundant components. Major peaks were observed corresponding to reported masses of IgG (~144 kD), albumin (~66 kD) , doubly charged albumin (~135 kD), transthyretin (~14 kD), apolipoprotein C-II(~8.9kD), apolipoprotein C-I (~6.7kd) and lysozyme (~14 kD). The measured mass values serve as a database for the expected positions for a number of plasma and CSF components when analyzed by MALDI-ToF MS. These data may be of value in helping to identify which proteins are forming peaks and in identifying the peaks that may be used as internal calibration masses in MS analysis of plasma and CSF. However, masses of some proteins may vary owing to sequence polymorphism, variable posttranslational modification, or degradation during processing or storage. Measured masses of highly sialyted proteins also may vary. It must also be considered that there is structural variation of proteins in different physiological states that may lead to changes in glycoprotein mass (Higai et al., 2005).

Direct analysis of diluted plasma and CSF by MALDI-ToF MS may offer a direct approach to analyze several major components with minimal modification or selective loss of components. For proteins that are not modified by post-translational modifications, the mass measurements determined by MALDI-ToF MS should be accurate measures of protein mass. However, the sensitivity of MALDI-ToF protein profiling remains poor when compared with antibody-based methods, as previously pointed out by Diamandis (Diamandis, 2004b, Diamandis, 2004a). Furthermore, protein profiling experiments of

unfractionated plasma and CSF are restricted to the limited mass range for protein detection with MALDI ToF MS (most ion signals between *m/z* 2,000 and *m/z* 30,000). This is due to the fact that dominance of the albumin peaks for monomer, doubly charged monomer and singly charged monomer interfere with the ability to detect other peaks with m/z > 30,000. Future analysis will include use of other analytic methods, including LC fractionation prior to MS as well as profiling enzymatic digests of plasma proteins, which not only increase the sensitivity and widen the accessible mass range for the protein detection, but also enable direct identification using tandem mass spectrometry peptide sequencing. To add value to the data, analysis of purified plasma and CSF proteins would help further identify the proteins corresponding to the mass values and also identify any modifications that may occur during sample preparation.

6.5 Conclusions

Some information about profile variation of major proteins in plasma and CSF may be obtained by the simple technique described here. As presented here, the analysis is primarily qualitative in nature. There are major issues with the results presented here and this study has just made a start but needs significant investment to see if this is a solution to CM biomarker discovery. Any effort to apply this as a quantitative technique for profiling the relative amounts of different proteins would require extensive investigation of appropriate dilutions, the linearity, and calibration of responses and standardisation procedures. Qualitative analysis of the global pattern may have value for a number of purposes. It may assist in evaluating the degree of fractionation achieved by other techniques such as SELDI-ToF MS. However, available algorithms for analysing both

SELDI-ToF and MALDI-ToF data have had problems when other laboratories have tried to validate them.

Nevertheless, owing to its simplicity, the approach described here may be of clinical value during the acute phase response, where analysis of abundant components provides valuable information. However, if profiles of low abundant proteins are to be created, there needs to be an extensive effort at specimen fractionation.

7 General Discussion and Conclusions

7.1 Introduction

Proteome alterations in disease may occur in many different ways that are not predictable from genomic analysis, and it is clear that a better understanding of these alterations will have a substantial impact in medicine. A useful repertoire of proteomics technologies is currently available for disease-related applications, although further technological innovations would be beneficial to increase sensitivity, reduce sample requirement, increase throughput and more effectively and uncover various types of protein alterations such as post-translational modifications. The use of these technologies will likely expand substantially, particularly to meet the need for better diagnostics and to shorten the path for developing effective therapy.

The overall aim of the studies described in this thesis was to analyse the global proteome of CSF and plasma collected from children diagnosed with CM. Samples were first analysed using 2DE followed by mass spectrometry and a gel free alternative using LC MS/MS. The latter method employed orthogonal liquid chromatography steps in conjunction with automated tandem mass spectrometric analysis. The final study described in the thesis was the creation of a mass profile of rare proteins associated with CM using MALDI-ToF analysis.

7.2 2-DE Gel Analysis of Plasma and CSF

Over the years 2-DE has proven to be a reliable and efficient method for separation of several hundred to a few thousand proteins on the basis of differences in their isoelectric

point (pI) and molecular weight. Although gel electrophoresis has many advantages, there are also several disadvantages. The procedure is long, labour-intensive and lacks sufficient reproducibility. The insolubility of hydrophobic proteins and the difficulty in detection and separation of LAPs are additional weaknesses.

In the studies described in chapters 3 and 4 the dynamic range of the technique limited analysis to the more abundant acute-phase proteins, including serum albumin and various isoforms of apolipoproteins, which are known markers of inflammatory disease. Achieving the level of detection commensurate with immunoassay will only be possible on 2-DE using pre-fractionated plasma/CSF in which the major proteins (albumin and immunoglobulins) are substantially reduced, but not at the expense of the nonspecific depletion of minor proteins. When this occurs, it will be possible to identify many subtle changes in protein expression, permitting protein enrichment to determine the mass spectrum and hence identity prior to development of specific immunoassays.

7.3 LC MS/MS of Plasma and CSF

In this thesis the use of LC-MS/MS as an alternative to 2-DE and a comparison of various separation techniques were explored. RP-LC has been used in many studies to overcome the shortcomings of 2-DE. However, the same problem of identifying the abundant proteins only is not overcome and in chapter 5 the use of a pre-fractionating method prior to multi-dimensional HPLC was described. This alternative method helped reduce the complexity of the sample being introduced into the MS. This technique enabled the identification of some LAPs with some measure of reproducibility. However, further validation with more samples would have to be done to ensure that the proteins are present with every sample run. The large number of proteins identified would also need to be

confirmed using techniques such as western blots to substantiate any further use of the expensive tool.

Despite improvements made, the proteins identified still did not account for even a third of those reported in literature and there were still a number of small proteins including several cytokines and interleukins implicated in malarial infection that were not identified. This may be due to the fact that some of the LAPs may have broken down or been modified due to long storage as well as due to how the sample was initially collected. This can only be done by using freshly collected samples and testing whether heparin and EDTA affect proteomic results of CSF and plasma.

7.4 Mass Profiling of Plasma and CSF

MALDI-ToF protein profiling has been applied in proteomics biomarker research in our study as well as in the study described in chapter 6; however, several aspects of MALDI protein profiling need further evaluation and optimisation before clinical use (Hortin, 2005). First, high-throughput MALDI protein profiling is presently confined to the detection of highly abundant proteins (Hortin, 2006). Sensitivity in clinical proteomics can be effectively improved with extensive prefractionation strategies, which still need critical evaluation before they can be used in high-throughput protein profiling. In one study, the use of reversed-phase HPLC and MALDI protein profiling to analyze plasma samples in a medium-throughput setup led to improved sensitivity (Tammen et al., 2005). Second, MALDI protein profiling provides only a limited mass window. The matrix molecules produce highly intense signals in the low-mass range (~0–1000 *Mr*) that obscure some of the signals of small peptides, and larger proteins have a lower detection limit, primarily because they fly slower than smaller proteins and detector response is related to ion

velocity (Hortin, 2006). The 3rd challenge in MALDI protein profiling is establishing the reproducibility of peak intensity. In biomarker research with MALDI protein profiling, the aim is to identify peak intensities (or peak areas) that are different between case and control samples, and the reproducibility of peak intensities is of highest importance. However, poor reproducibility has been considered one of the major problems in protein profiling with MALDI-ToF MS. The matrix (co)crystallization and desorption/ionization steps in MALDI-ToF MS have been derived empirically, and the processes are poorly understood. Different matrix molecules crystallize in different shapes and dimensions, proteins tend to accumulate at the droplet periphery, and the composition of the matrix solution and the rate of crystal growth influence the spectral output (Cohen and Chait, 1996). These phenomena produce shot-to-shot variation, which are related to sampling different parts of the target surface and progressive sample ablation with repeated sampling. The desorption/ionization step in MALDI-ToF MS is a complex process involving optical and mechanical phenomena, as well as thermodynamic and physicochemical processes of phase transition and ionization, which are not well understood (Dreisewerd, 2003). Studies have demonstrated ion suppression effects in MALDI-ToF MS. Ion suppression occurs when an ion suppresses the peak signal of other ions in the sample, and peptides with greater hydrophobicity show the greatest suppression effects. The presence of basic residues may favour ionisation in MALDI-ToF MS analysis (Bohring et al., 2001). One study found that highly acidic compounds produced weak signals in MALDI-ToF MS analyses, but when such compounds were mixed with a basic peptide to form a noncovalent complex, the signals improved (Juhasz and Biemann, 1994). In summary, peak intensity in MALDI protein profiling has significant analytical variation and is poorly understood. Peak intensity is related to the concentration of the individual

protein, to its primary structure, and to the complexity of the sample. To enhance quantitive capability of protein profiling it will be necessary to develop better peak isolation methods.

Masses found to be unique to the disease groups are generally attributed to common acute phase proteins. To further enhance the peak capability in CM biomarker studies extensive studies using samples that have been fractionated will have to be done. Using a larger sample size to validate any spectra will also be done.

7.5 Suggested Future Studies

The results presented in this thesis present a basis for further studies aimed at utilising proteomic tools to better understand the pathophysiology of severe disease in the hope of identifying new biomarkers. The following is an outline of studies that could be conducted:

a) A prospective study to confirm that any of the differences in protein profiles are not due to long-term storage of samples
b) Use of prefractionation steps to improve mass profiles created using MALDI-ToF.
c) The use of newly developed quantitative proteomic techniques such as isotope-coded affinity tags (ICAT), iTRAQ and DIGE to improve quantification of differentially expressed proteins in severe disease.

7.6 Conclusion

In so far as biological implications, the results in this thesis partly support the hypothesis that CM is an inflammatory like process because some of the regulated proteins are physiologically linked to the control of inflammation. However, the difference between

proteins identified in ABM and CM would indicate a difference in the overall inflammation process between the two groups.

BBB permeability in CM is still described as moderate despite the major neurological complications associated with it and results on total protein content in CSF supports this claim. One way of testing the level of breakdown is by assessing the albumin and IgG CSF/serum ratios. In the thesis other acute phase proteins such as transferrin seem to be more abundantly available than IgG, in both CSF and plasma and may prove to give a better ratio in trying to understand the mechanism behind BBB permeability in CM. Another protein identified in both the CSF and plasma is human Spectrin beta chain brain 3 protein and it would be of interest to see whether there are self reactive antibodies to this protein as in the Gabon study described by Guiyedi *et. al* on the alpha chain of the same protein and whether these antibodies are as a consequence to parasite ex posure, and whether their levels might parallel the degree of neuronal damage.

Work in this thesis has laid groundwork for establishing the proteomic component of eventual systems biology of severe malaria as well as other severe tropical diseases such as HIV. Of interest to immunologists, is whether the repertoire of parasite proteins identified in this study plays a role in immunity against malaria. However, verification of the parasite proteins identified would need to be done before any conclusions can be made.

In summary, work on the aspects of proteomics studied has generated renewed confidence that such approaches will reveal important features of normal biology and physiology and assist in the discovery, validation, and application of protein biomarker panels in early diagnosis of disease and monitoring responses to therapies. However, before the full

impact of these technologies can be realised, some technical hurdles must be resolved. This can be done by combining novel protein separation technologies and mass spectrometry approaches to improve the detection and identification of low abundance proteins. New technology platforms for global analysis using LTQ-FT, LC-MS/MS/MS, and LTQ-Orbitrap (Olsen et al., 2004) and new platforms for targeted proteomics using heavy-isotope- labeled N-glycosite-containing proteotypic peptides (Kuster et al., 2005, Aebersold and Mann, 2003) and/or multiple reaction monitoring (MRM) with or without anti-peptide antibodies (Polanski and Anderson, 2006) offer great promise. Additional mining of high-quality spectra may increase yields of protein identifications, as well. As mentioned earlier methods of quantitative proteomics, such as ICAT, iTRAQ, and DIGE may be essential. Aebersold *et al* are developing a resource that will offer chemically synthesized peptides tagged with heavy isotope for each gene and eventually each protein isoform needed for high discrimination as biomarkers (Aebersold, 2003). These "proteotypic peptides" would permit spiking of specimens to facilitate identification of mass pairs with the same peptides in the biological specimen and quantitation of the peptide and its protein (Kuster et al., 2005). Another approach could be the use of sensitive anti-peptide antibody-enhanced assays for quantitation lower-abundance proteins (Anderson and Hunter, 2006).

Technological improvements will have an impact on the ability to discover novel and important markers. The development of a fast and cost-effective mass spectrometry based assay, will greatly improve the ability to validate (or invalidate) promising new markers and will eventually lead to a credible application of protein biomarkers in population screening and eventually in patient care.

ABDI, F., QUINN, J. F., JANKOVIC, J., MCINTOSH, M., LEVERENZ, J. B., PESKIND, E., NIXON, R., NUTT, J., CHUNG, K., ZABETIAN, C., SAMII, A., LIN, M., HATTAN, S., PAN, C., WANG, Y., JIN, J., ZHU, D., LI, G. J., LIU, Y., WAICHUNAS, D., MONTINE, T. J. & ZHANG, J. (2006) Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. *J Alzheimers Dis,* 9, 293-348.

ADAM, B. L., QU, Y., DAVIS, J. W., WARD, M. D., CLEMENTS, M. A., CAZARES, L. H., SEMMES, O. J., SCHELLHAMMER, P. F., YASUI, Y., FENG, Z. & WRIGHT, G. L., JR. (2002) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res,* 62, 3609-14.

ADAMS, S., BROWN, H. & TURNER, G. (2002) Breaking down the blood-brain barrier: signaling a path to cerebral malaria? *Trends Parasitol,* 18, 360-6.

AEBERSOLD, R. (2003) Constellations in a cellular universe. *Nature,* 422, 115-6.

AEBERSOLD, R. & MANN, M. (2003) Mass spectrometry-based proteomics. *Nature,* 422, 198-207.

AHMED, N., BARKER, G., OLIVA, K., GARFIN, D., TALMADGE, K., GEORGIOU, H., QUINN, M. & RICE, G. (2003) An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. *Proteomics,* 3, 1980-7.

ALAIYA, A. A., FRANZEN, B., AUER, G. & LINDER, S. (2000) Cancer proteomics: from identification of novel markers to creation of artifical learning models for tumor classification. *Electrophoresis,* 21, 1210-7.

ALEXE, G., ALEXE, S., LIOTTA, L. A., PETRICOIN, E., REISS, M. & HAMMER, P. L. (2004) Ovarian cancer detection by logical analysis of proteomic data. *Proteomics,* 4, 766-83.

ALTELAAR, A. F., KLINKERT, I., JALINK, K., DE LANGE, R. P., ADAN, R. A., HEEREN, R. M. & PIERSMA, S. R. (2006) Gold-enhanced biomolecular surface imaging of cells and tissue by SIMS and MALDI mass spectrometry. *Anal Chem,* 78, 734-42.

AMES, G. F. & NIKAIDO, K. (1976) Two-dimensional gel electrophoresis of membrane proteins. *Biochemistry,* 15, 616-23.

ANDERSON, L. (2005) Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J Physiol,* 563, 23-60.

ANDERSON, L. & HUNTER, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics,* 5, 573-88.

ANDERSON, N. L. & ANDERSON, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics,* 1, 845-67.

ANDERSON, N. L., POLANSKI, M., PIEPER, R., GATLIN, T., TIRUMALAI, R. S., CONRADS, T. P., VEENSTRA, T. D., ADKINS, J. N., POUNDS, J. G., FAGAN,

R. & LOBLEY, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics,* 3, 311-26.

ATKINSON, A. J., COLBURN, W. A., DEGRUTTOLA, V. G., DEMETS, D. L., DOWNING, G. J., HOTH, D. F., OATES, J. A., PECK, C. C., SCHOOLEY, R. T., SPILKER, B. A., WOODCOCK, J. & ZEGER, S. L. (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics.*

BAGGERLY, K. A., MORRIS, J. S. & COOMBES, K. R. (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics,* 20, 777-785.

BAGGERLY, K. A., MORRIS, J. S., EDMONSON, S. R. & COOMBES, K. R. (2005) Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Natl Cancer Inst,* 97, 307-9.

BAIROCH, A. & APWEILER, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res,* 28, 45-8.

BANEZ, L. L., PRASANNA, P., SUN, L., ALI, A., ZOU, Z., ADAM, B. L., MCLEOD, D. G., MOUL, J. W. & SRIVASTAVA, S. (2003) Diagnostic potential of serum proteomic patterns in prostate cancer. *J Urol,* 170, 442-6.

BANKS, R. E., DUNN, M. J., HOCHSTRASSER, D. F., SANCHEZ, J. C., BLACKSTOCK, W., PAPPIN, D. J. & SELBY, P. J. (2000) Proteomics: new perspectives, new biomedical opportunities. *Lancet,* 356, 1749-56.

BARKER, W. C., GARAVELLI, J. S., HUANG, H., MCGARVEY, P. B., ORCUTT, B. C., SRINIVASARAO, G. Y., XIAO, C., YEH, L. S., LEDLEY, R. S., JANDA, J. F., PFEIFFER, F., MEWES, H. W., TSUGITA, A. & WU, C. (2000) The protein information resource (PIR). *Nucleic Acids Res,* 28, 41-4.

BARON, C. & THOMPSON, T. E. (1975) Solubilization of bacterial membrane proteins using alkyl glucosides and dioctanoyl phosphatidylcholine. *Biochim Biophys Acta,* 382, 276-85.

BARRETT, J., JEFFERIES, J. R. & BROPHY, P. M. (2000) Parasite proteomics. *Parasitol Today,* 16, 400-3.

BEAVIS, R. C. & CHAIT, B. T. (1990) High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry. *Anal Chem,* 62, 1836-40.

BEESLEY, R., FILTEAU, S., TOMKINS, A., DOHERTY, T., AYLES, H., REID, A., ELLMAN, T. & PARTON, S. (2000) Impact of acute malaria on plasma concentrations of transferrin receptors. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 94, 295.

BEESON, J. G. & BROWN, G. V. (2002) Pathogenesis of Plasmodium falciparum malaria: the roles of parasite adhesion and antigenic variation. *Cell Mol Life Sci,* 59, 258-71.

BELGACEM, O., BUCHACHER, A., POCK, K., JOSIC, D., SUTTON, C., RIZZI, A. & ALLMAIER, G. (2002) Molecular mass determination of plasma-derived glycoproteins by ultraviolet matrix-assisted laser desorption/ionization time-of-flight mass spectrometry with internal calibration. *J Mass Spectrom,* 37, 1118-30.

BERANOVA-GIORGIANNI, S. & DESIDERIO, D. M. (2000) Mass spectrometry of the human pituitary proteome: identification of selected proteins. *Rapid Commun Mass Spectrom,* 14, 161-7.

BERKELMAN, T. R. & LAGARIAS, J. C. (1986) Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal Biochem,* 156, 194-201.

BERKLEY, J. A., MWANGI, I., MELLINGTON, F., MWARUMBA, S. & MARSH, K. (1999) Cerebral malaria versus bacterial meningitis in children with impaired consciousness. *Qjm,* 92, 151-7.

BICHSEL, V. E., LIOTTA, L. A. & PETRICOIN, E. F., 3RD (2001) Cancer proteomics: from biomarker discovery to signal pathway profiling. *Cancer J,* 7, 69-78.

BIEMANN, K. (1988) Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom,* 16, 99-111.

BJELLQVIST, B., EK, K., RIGHETTI, P. G., GIANAZZA, E., GORG, A., WESTERMEIER, R. & POSTEL, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods,* 6, 317-39.

BJORHALL, K., MILIOTIS, T. & DAVIDSSON, P. (2005) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics,* 5, 307-17.

BLUM, H. E., HAASE, A. T. & VYAS, G. N. (1984) Molecular pathogenesis of hepatitis B virus infection: simultaneous detection of viral DNA and antigens in paraffin-embedded liver sections. *Lancet,* 2, 771-5.

BOHRING, C., KRAUSE, E., HABERMANN, B. & KRAUSE, W. (2001) Isolation and identification of sperm membrane antigens recognized by antisperm antibodies, and their possible role in immunological infertility disease. *Mol Hum Reprod,* 7, 113-8.

BRADFORD, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem,* 72, 248-54.

BRIVIO, M., FOKKENS, R. H., VERBOOM, W., REINHOUDT, D. N., TAS, N. R., GOEDBLOED, M. & VAN DEN BERG, A. (2002) Integrated microfluidic system enabling (bio)chemical reactions with on-line MALDI-TOF mass spectrometry. *Anal Chem,* 74, 3972-6.

BROWN, H., HIEN, T. T., DAY, N., MAI, N. T., CHUONG, L. V., CHAU, T. T., LOC, P. P., PHU, N. H., BETHELL, D., FARRAR, J., GATTER, K., WHITE, N. & TURNER, G. (1999) Evidence of blood-brain barrier dysfunction in human cerebral malaria. *Neuropathol Appl Neurobiol,* 25, 331-40.

BRUCE-CHWATT, L. (1988) History of Malaria from Prehsitory to Eradication in Malaria. IN WERNSDORFER, W. H. & I., M. (Eds.) *Malaria: Principles and Practice of Malariology.* Edinburgh, Churchill Livingstone.

BRUSH, C. F., WALTER, M. F., ANGUISH, L. J. & GHIORSE, W. C. (1998) Influence of pretreatment and experimental conditions on electrophoretic mobility and hydrophobicity of Cryptosporidium parvum oocysts. *Appl Environ Microbiol,* 64, 4439-45.

BURGESS, R. R. & THOMPSON, N. E. (2002) Advances in gentle immunoaffinity chromatography. *Curr Opin Biotechnol,* 13, 304-8.

BUTT, A., DAVISON, M. D., SMITH, G. J., YOUNG, J. A., GASKELL, S. J., OLIVER, S. G. & BEYNON, R. J. (2001) Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis. *Proteomics,* 1, 42-53.

CAPRIOLI, R. M., FARMER, T. B. & GILE, J. (1997) Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal Chem,* 69, 4751-60.

CARLSON, J., HELMBY, H., HILL, A. V., BREWSTER, D., GREENWOOD, B. M. & WAHLGREN, M. (1990a) Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet,* 336, 1457-60.

CARLSON, J., HOLMQUIST, G., TAYLOR, D. W., PERLMANN, P. & WAHLGREN, M. (1990b) Antibodies to a histidine-rich protein (PfHRP1) disrupt spontaneously formed Plasmodium falciparum erythrocyte rosettes. *Proc Natl Acad Sci U S A,* 87, 2511-5.

CARR, S., AEBERSOLD, R., BALDWIN, M., BURLINGAME, A., CLAUSER, K. & NESVIZHSKII, A. (2004) The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics,* 3, 531-3.

CATINELLA, S., SERAGLIA, R. & MARSILIO, R. (1999) Evaluation of protein profile of human milk by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom,* 13, 1546-9.

CHAGA, G. S. (2001) Twenty-five years of immobilized metal ion affinity chromatography: past, present and future. *Journal of Biochemical and Biophysical Methods,* 49, 313.

CHAIT, B. T. & KENT, S. B. (1992) Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science,* 257, 1885-94.

CHANG, W. L., JONES, S. P., LEFER, D. J., WELBOURNE, T., SUN, G., YIN, L., SUZUKI, H., HUANG, J., GRANGER, D. N. & VAN DER HEYDE, H. C. (2001) CD8(+)-T-cell depletion ameliorates circulatory shock in Plasmodium berghei-infected mice. *Infect Immun,* 69, 7341-8.

CHANOINE, J. P., ALEX, S., FANG, S. L., STONE, S., LEONARD, J. L., KORHLE, J. & BRAVERMAN, L. E. (1992) Role of transthyretin in the transport of thyroxine from the blood to the choroid plexus, the cerebrospinal fluid, and the brain. *Endocrinology,* 130, 933-8.

CLARK, I. A., COWDEN, W. B. & ROCKETT, K. A. (1993) Nitric oxide and cerebral malaria. *Lancet,* 341, 632-3.

CLARK, I. A., ROCKETT, K. A. & COWDEN, W. B. (1991) Role of TNF in cerebral malaria. *Lancet,* 337, 302-3.

COHEN, S. L. & CHAIT, B. T. (1996) Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal Chem,* 68, 31-7.

COMBES, V., COLTEL, N., FAILLE, D., WASSMER, S. C. & GRAU, G. E. (2006) Cerebral malaria: role of microparticles and platelets in alterations of the blood-brain barrier. *Int J Parasitol,* 36, 541-6.

COMBES, V., ROSENKRANZ, A. R., REDARD, M., PIZZOLATO, G., LEPIDI, H., VESTWEBER, D., MAYADAS, T. N. & GRAU, G. E. (2004) Pathogenic role of P-selectin in experimental cerebral malaria: importance of the endothelial compartment. *Am J Pathol,* 164, 781-6.

COMPTON, S. J. & JONES, C. G. (1985) Mechanism of dye response and interference in the Bradford protein assay. *Anal Biochem,* 151, 369-74.

COSTELLO, M. J., GEWURZ, H. & SIEGEL, J. N. (1984) Inhibition of neutrophil activation by alpha1-acid glycoprotein. *Clin Exp Immunol,* 55, 465-72.

COWMAN, A. F. (2001) Functional analysis of drug resistance in Plasmodium falciparum in the post-genomic era. *Int J Parasitol,* 31, 871-8.

CROSSMAN, L., MCHUGH, N. A., HSIEH, Y., KORFMACHER, W. A. & CHEN, J. (2006) Investigation of the profiling depth in matrix-assisted laser desorption/ionization imaging mass spectrometry. *Rapid Commun Mass Spectrom,* 20, 284-90.

DAS, B. S., THURNHAM, D. I. & DAS, D. B. (1996) Plasma alpha-tocopherol, retinol, and carotenoids in children with falciparum malaria. *Am J Clin Nutr,* 64, 94-100.

DAVIDSSON, P., SJOGREN, M., ANDREASEN, N., LINDBJER, M., NILSSON, C. L., WESTMAN-BRINKMALM, A. & BLENNOW, K. (2002a) Studies of the pathophysiological mechanisms in frontotemporal dementia by proteome analysis of CSF proteins. *Brain Res Mol Brain Res,* 109, 128-33.

DAVIDSSON, P., WESTMAN-BRINKMALM, A., NILSSON, C. L., LINDBJER, M., PAULSON, L., ANDREASEN, N., SJOGREN, M. & BLENNOW, K. (2002b) Proteome analysis of cerebrospinal fluid proteins in Alzheimer patients. *Neuroreport,* 13, 611-5.

DE KOSSODO, S. & GRAU, G. E. (1993) Role of cytokines and adhesion molecules in malaria immunopathology. *Stem Cells,* 11, 41-8.

DIAMANDIS, E. P. (2003) Re: Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst,* 95, 489-90; author reply 490-1.

DIAMANDIS, E. P. (2004a) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst,* 96, 353-6.

DIAMANDIS, E. P. (2004b) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics,* 3, 367-78.

DIAMANDIS, E. P. (2004c) Proteomic patterns to identify ovarian cancer: 3 years on. *Expert Rev Mol Diagn,* 4, 575-7.

DIAMANDIS, E. P. (2006) Serum proteomic profiling by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for cancer diagnosis: next steps. *Cancer Res,* 66, 5540-1.

DICKSON, P. W., ALDRED, A. R., MENTING, J. G., MARLEY, P. D., SAWYER, W. H. & SCHREIBER, G. (1987) Thyroxine transport in choroid plexus. *J Biol Chem,* 262, 13907-15.

DREISEWERD, K. (2003) The desorption process in MALDI. *Chem Rev,* 103, 395-426.

DUDGEON, L. & CLARK, C. (1917) A Contribution to the Microscopial Histology of Malaria. *Lancet,* 2, 153-156.

ECHAN, L. A., TANG, H. Y., ALI-KHAN, N., LEE, K. & SPEICHER, D. W. (2005) Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics,* 5, 3292-303.

ENG, J. K., MCCORMACK, A. L. & YATES, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry,* 5, 976.

ENGIDAWORK, E. & LUBEC, G. (2001) Protein expression in Down syndrome brain. *Amino Acids,* 21, 331-61.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246, 64-71.

FICARRO, S. B., MCCLELAND, M. L., STUKENBERG, P. T., BURKE, D. J., ROSS, M. M., SHABANOWITZ, J., HUNT, D. F. & WHITE, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol,* 20, 301-5.

FILTEAU, S. M., MORRIS, S. S., ABBOTT, R. A., TOMKINS, A. M., KIRKWOOD, B. R., ARTHUR, P., ROSS, D. A., GYAPONG, J. O. & RAYNES, J. G. (1993) Influence of morbidity on serum retinol of children in a community-based study in northern Ghana. *Am J Clin Nutr,* 58, 192-7.

FOUNTOULAKIS, M. (2001) Proteomics: current technologies and applications in neurological disorders and toxicology. *Amino Acids,* 21, 363-81.

FOUNTOULAKIS, M. & LAHM, H. W. (1998) Hydrolysis and amino acid composition of proteins. *J Chromatogr A,* 826, 109-34.

FRIIS, H., MWANIKI, D., OMONDI, B., MUNIU, E., MAGNUSSEN, P., GEISSLER, W., THIONG'O, F. & MICHAELSEN, K. F. (1997) Serum retinol concentrations and Schistosoma mansoni, intestinal helminths, and malarial parasitemia: a cross-sectional study in Kenyan preschool and primary school children. *Am J Clin Nutr,* 66, 665-71.

FUKUYAMA, R., MIZUNO, T., MORI, S., YANAGISAWA, K., NAKAJIMA, K. & FUSHIKI, S. (2000) Age-dependent decline in the apolipoprotein E level in cerebrospinal fluid from control subjects and its increase in cerebrospinal fluid from patients with Alzheimer's disease. *Eur Neurol,* 43, 161-9.

FUTCHER, B., LATTER, G. I., MONARDO, P., MCLAUGHLIN, C. S. & GARRELS, J. I. (1999) A sampling of the yeast proteome. *Mol Cell Biol,* 19, 7357-68.

GABERC-POREKAR, V. & MENART, V. (2001) Perspectives of immobilized-metal affinity chromatography. *Journal of Biochemical and Biophysical Methods,* 49, 335.

GARNHAM, P. C. (1988a) History of discoveries of malaria parasites and of their life cycles. *Hist Philos Life Sci,* 10, 93-108.

GARNHAM, P. C. C. (1988b) Malaria parasites of man: life cycles and morphology (excluding untrastructure). IN WERNSDORFER, W. H. & MCGREGOR, I. A. (Eds.) *Malaria: principles and practice of malariology.* Edinburgh, Churchill Livingstone.

GARRETT, T. J. & YOST, R. A. (2006) Analysis of intact tissue by intermediate-pressure MALDI on a linear ion trap mass spectrometer. *Anal Chem,* 78, 2465-9.

GITAU, E. N. & NEWTON, C. R. (2005) Review Article: Blood-brain barrier in falciparum malaria*. *Trop Med Int Health,* 10, 285-92.

GORG, A., BOGUTH, G., OBERMAIER, C., POSCH, A. & WEISS, W. (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): the state of the art and the controversy of vertical versus horizontal systems. *Electrophoresis,* 16, 1079-86.

GORG, A., BOGUTH, G., OBERMAIER, C. & WEISS, W. (1998) Two-dimensional electrophoresis of proteins in an immobilized pH 4-12 gradient. *Electrophoresis,* 19, 1516-9.

GORG, A., OBERMAIER, C., BOGUTH, G., CSORDAS, A., DIAZ, J. J. & MADJAR, J. J. (1997) Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis,* 18, 328-37.

GORG, A., OBERMAIER, C., BOGUTH, G., HARDER, A., SCHEIBE, B., WILDGRUBER, R. & WEISS, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis,* 21, 1037-53.

GORG, A. & WEISS, W. (1999) Horizontal SDS-PAGE for IPG-Dalt. *Methods Mol Biol,* 112, 235-44.

GORG, A., WEISS, W. & DUNN, M. J. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics,* 4, 3665-85.

GRANGER, J., SIDDIQUI, J., COPELAND, S. & REMICK, D. (2005) Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics,* 5, 4713-8.

GRANZIER, H. L. & WANG, K. (1993) Gel electrophoresis of giant proteins: solubilization and silver-staining of titin and nebulin from single muscle fiber segments. *Electrophoresis,* 14, 56-64.

GRAU, G. E., FAJARDO, L. F., PIGUET, P. F., ALLET, B., LAMBERT, P. H. & VASSALLI, P. (1987) Tumor necrosis factor (cachectin) as an essential mediator in murine cerebral malaria. *Science,* 237, 1210-2.

GRAU, G. E., FREI, K., PIGUET, P. F., FONTANA, A., HEREMANS, H., BILLIAU, A., VASSALLI, P. & LAMBERT, P. H. (1990) Interleukin 6 production in experimental cerebral malaria: modulation by anticytokine antibodies and possible role in hypergammaglobulinemia. *J Exp Med,* 172, 1505-8.

GRAU, G. E., KINDLER, V., PIGUET, P. F., LAMBERT, P. H. & VASSALLI, P. (1988) Prevention of experimental cerebral malaria by anticytokine antibodies. Interleukin 3 and granulocyte macrophage colony-stimulating factor are intermediates in increased tumor necrosis factor production and macrophage accumulation. *J Exp Med,* 168, 1499-504.

GRAU, G. E. & LOU, J. N. (1995) Experimental cerebral malaria: possible new mechanisms in the TNF-induced microvascular pathology. *Soz Praventivmed,* 40, 50-7.

GRAU, G. E., PIGUET, P. F., VASSALLI, P. & LAMBERT, P. H. (1989) Tumor-necrosis factor and other cytokines in cerebral malaria: experimental and clinical data. *Immunol Rev,* 112, 49-70.

GUHA, M., CHOUBEY, V., MAITY, P., KUMAR, S., SHRIVASTAVA, K., PURI, S. K. & BANDYOPADHYAY, U. (2007) Overexpression, purification and localization of apoptosis related protein from Plasmodium falciparum. *Protein Expr Purif,* 52, 363-72.

GUHA, M., KUMAR, S., CHOUBEY, V., MAITY, P. & BANDYOPADHYAY, U. (2006) Apoptosis in liver during malaria: role of oxidative stress and implication of mitochondrial pathway. *Faseb J,* 20, 1224-6.

GUIYEDI, V., CHANSEAUD, Y., FESEL, C., SNOUNOU, G., ROUSSELLE, J. C., LIM, P., KOKO, J., NAMANE, A., CAZENAVE, P. A., KOMBILA, M. & PIED, S.

(2007) Self-reactivities to the non-erythroid alpha spectrin correlate with cerebral malaria in gabonese children. *PLoS ONE,* 2, e389.

GYGI, S. P. & AEBERSOLD, R. (2000) Mass spectrometry and proteomics. *Curr Opin Chem Biol,* 4, 489-94.

GYGI, S. P., CORTHALS, G. L., ZHANG, Y., ROCHON, Y. & AEBERSOLD, R. (2000a) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A,* 97, 9390-5.

GYGI, S. P., RIST, B. & AEBERSOLD, R. (2000b) Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol,* 11, 396-401.

GYGI, S. P., RIST, B., GERBER, S. A., TURECEK, F., GELB, M. H. & AEBERSOLD, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol,* 17, 994-9.

HAMES, B. D. (1994) *Gel Electrophoresis of proteins,* Oxford, Oxford Unversity Press.

HAMMACK, B. N., OWENS, G. P., BURGOON, M. P. & GILDEN, D. H. (2003) Improved resolution of human cerebrospinal fluid proteins on two-dimensional gels. *Mult Scler,* 9, 472-5.

HANASH, S. (2003) Disease proteomics. *Nature,* 422, 226-32.

HANASH, S. M., MADOZ-GURPIDE, J. & MISEK, D. E. (2002) Identification of novel targets for cancer therapy using expression proteomics. *Leukemia,* 16, 478-85.

HARRISON, P. A. & BLACK, C. C. (1982) Two-Dimensional Electrophoretic Mapping of Proteins of Bundle Sheath and Mesophyll Cells of the C(4) Grass Digitaria sanguinalis (L.) Scop. (Crabgrass). *Plant Physiol,* 70, 1359-1366.

HAYNES, P. A. & YATES, J. R., 3RD (2000) Proteome profiling-pitfalls and progress. *Yeast,* 17, 81-7.

HEARN, J., RAYMENT, N., LANDON, D. N., KATZ, D. R. & DE SOUZA, J. B. (2000) Immunopathology of cerebral malaria: morphological evidence of parasite sequestration in murine brain microvasculature. *Infect Immun,* 68, 5364-76.

HEDDINI, A., PETTERSSON, F., KAI, O., SHAFI, J., OBIERO, J., CHEN, Q., BARRAGAN, A., WAHLGREN, M. & MARSH, K. (2001) Fresh isolates from children with severe Plasmodium falciparum malaria bind to multiple receptors. *Infect Immun,* 69, 5849-56.

HERBERT, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis,* 20, 660-3.

HERBERT, B. R., MOLLOY, M. P., GOOLEY, A. A., WALSH, B. J., BRYSON, W. G. & WILLIAMS, K. L. (1998) Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. *Electrophoresis,* 19, 845-51.

HERTZ-FOWLER, C., PEACOCK, C. S., WOOD, V., ASLETT, M., KERHORNOU, A., MOONEY, P., TIVEY, A., BERRIMAN, M., HALL, N., RUTHERFORD, K., PARKHILL, J., IVENS, A. C., RAJANDREAM, M.-A. & BARRELL, B. (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucl. Acids Res.,* 32, D339-343.

HIGAI, K., AOKI, Y., AZUMA, Y. & MATSUMOTO, K. (2005) Glycosylation of site-specific glycans of alpha1-acid glycoprotein and alterations in acute and chronic inflammation. *Biochim Biophys Acta,* 1725, 128-35.

HILLENKAMP, F., KARAS, M., BEAVIS, R. C. & CHAIT, B. T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem,* 63, 1193A-1203A.

HOFFMANN, P., JI, H., MORITZ, R. L., CONNOLLY, L. M., FRECKLINGTON, D. F., LAYTON, M. J., EDDES, J. S. & SIMPSON, R. J. (2001) Continuous free-flow electrophoresis separation of cytosolic proteins from the human colon carcinoma cell line LIM 1215: a non two-dimensional gel electrophoresis-based proteome analysis strategy. *Proteomics,* 1, 807-18.

HORTIN, G. & REMALEY, A. (2006) Mass determination of major plasma proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clinical Proteomics,* 2, 103.

HORTIN, G. L. (2005) Can mass spectrometric protein profiling meet desired standards of clinical laboratory practice? *Clin Chem,* 51, 3-5.

HORTIN, G. L. (2006) The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin Chem,* 52, 1223-37.

HUANG, L., HARVIE, G., FEITELSON, J. S., GRAMATIKOFF, K., HEROLD, D. A., ALLEN, D. L., AMUNNGAMA, R., HAGLER, R. A., PISANO, M. R., ZHANG, W. W. & FANG, X. (2005) Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis. *Proteomics,* 5, 3314-28.

HUMPHERY-SMITH, I., CORDWELL, S. J. & BLACKSTOCK, W. P. (1997) Proteome research: complementarity and limitations with respect to the RNA and DNA worlds. *Electrophoresis,* 18, 1217-42.

IDRO, R., CARTER, J. A., FEGAN, G., NEVILLE, B. G. & NEWTON, C. R. (2006) Risk factors for persisting neurological and cognitive impairments following cerebral malaria. *Arch Dis Child,* 91, 142-8.

ISHIZUKA, N., KOBAYASHI, H., MINAKUCHI, H., NAKANISHI, K., HIRAO, K., HOSOYA, K., IKEGAMI, T. & TANAKA, N. (2002) Monolithic silica columns for high-efficiency separations by high-performance liquid chromatography. *J Chromatogr A,* 960, 85-96.

ISSAQ, H. J., CONRADS, T. P., PRIETO, D. A., TIRUMALAI, R. & VEENSTRA, T. D. (2003) SELDI-TOF MS for diagnostic proteomics. *Anal Chem,* 75, 148A-155A.

JAKKA, S. R., VEENA, S., ATMAKURI, R. M. & EISENHUT, M. (2006) Characteristic abnormalities in cerebrospinal fluid biochemistry in children with cerebral malaria compared to viral encephalitis. *Cerebrospinal Fluid Res,* 3, 8.

JENNINGS, V. M., ACTOR, J. K., LAL, A. A. & HUNTER, R. L. (1997) Cytokine profile suggesting that murine cerebral malaria is an encephalitis. *Infect Immun,* 65, 4883-7.

JING, Z. (2007) Proteomics of human cerebrospinal fluid - the good, the bad, and the ugly. *PROTEOMICS - CLINICAL APPLICATIONS,* 1, 805-819.

JUHASZ, P. & BIEMANN, K. (1994) Mass spectrometric molecular-weight determination of highly acidic compounds of biological significance via their complexes with basic polypeptides. *Proc Natl Acad Sci U S A,* 91, 4333-7.

KARAS, M., BACHMANN, D., BAHR, U. & HILLENKAMP, F. (1987) Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes,* 78, 53.

KARAS, M., GLUCKMANN, M. & SCHAFER, J. (2000) Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom,* 35, 1-12.

KENNEDY, S. (2001) Proteomic profiling from human samples: the body fluid alternative. *Toxicol Lett,* 120, 379-84.

KENYON, G. L., DEMARINI, D. M., FUCHS, E., GALAS, D. J., KIRSCH, J. F., LEYH, T. S., MOOS, W. H., PETSKO, G. A., RINGE, D., RUBIN, G. M. & SHEAHAN, L. C. (2002) Defining the Mandate of Proteomics in the Post-Genomics Era: Workshop Report: (C)2002 National Academy of Sciences, Washington, D.C., USA. Reprinted with permission from the National Academies Press for the National Academy of Sciences. All rights reserved. The original report may be viewed online at http://www.nap.edu/catalog/10209.html. *Mol Cell Proteomics,* 1, 763-780.

KETTMAN, J. R., FREY, J. R. & LEFKOVITS, I. (2001) Proteome, transcriptome and genome: top down or bottom up analysis? *Biomol Eng,* 18, 207-12.

KLOSE, J. & SPIELMANN, H. (1975) Gel isoelectric focusing of mouse lactate dehydrogenase: heterogeneity of the isoenzymes A4 and X4. *Biochem Genet,* 13, 707-20.

KOICHI TANAKA, H. W. Y. I. S. A. Y. Y. T. Y. T. M. (1988) Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2, 151-153.

KORENROMP, E. L. (2005) Malaria Incidence Estimates at Country Level for the Year 2004 – Proposed Estimates and Draft Reports

Geneva, Roll Back Malaria, World Health Organisation.

KUSTER, B., SCHIRLE, M., MALLICK, P. & AEBERSOLD, R. (2005) Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol,* 6, 577-83.

LAEMMLI, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227, 680-5.

LE BIHAN, T., DUEWEL, H. S. & FIGEYS, D. (2003) On-line strong cation exchange micro-HPLC-ESI-MS/MS for protein identification and process optimization. *J Am Soc Mass Spectrom,* 14, 719-27.

LEINWEBER, F. C. & TALLAREK, U. (2003) Chromatographic performance of monolithic and particulate stationary phases. Hydrodynamics and adsorption capacity. *J Chromatogr A,* 1006, 207-28.

LI, C. & LEE, K. H. (2004) Affinity depletion of albumin from human cerebrospinal fluid using Cibacron-blue-3G-A-derivatized photopatterned copolymer in a microfluidic device. *Anal Biochem,* 333, 381-8.

LI, J., LERICHE, T., TREMBLAY, T. L., WANG, C., BONNEIL, E., HARRISON, D. J. & THIBAULT, P. (2002a) Application of microfluidic devices to proteomics research: identification of trace-level protein digests and affinity capture of target peptides. *Mol Cell Proteomics,* 1, 157-68.

LI, J., ZHANG, Z., ROSENZWEIG, J., WANG, Y. Y. & CHAN, D. W. (2002b) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem,* 48, 1296-304.

LIM, H., ENG, J., YATES, J. R., 3RD, TOLLAKSEN, S. L., GIOMETTI, C. S., HOLDEN, J. F., ADAMS, M. W., REICH, C. I., OLSEN, G. J. & HAYS, L. G. (2003) Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide

mass mapping to mu LC-ESI tandem mass spectrometry. *J Am Soc Mass Spectrom,* 14, 957-70.

LINK, A. J., ENG, J., SCHIELTZ, D. M., CARMACK, E., MIZE, G. J., MORRIS, D. R., GARVIK, B. M. & YATES, J. R., 3RD (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol,* 17, 676-82.

LISTON, P., LEFEBVRE, C., FONG, W. G., XUAN, J. Y. & KORNELUK, R. G. (1997) Genomic characterization of the mouse inhibitor of apoptosis protein 1 and 2 genes. *Genomics,* 46, 495-503.

LIU, H., BERGER, S. J., CHAKRABORTY, A. B., PLUMB, R. S. & COHEN, S. A. (2002) Multidimensional chromatography coupled to electrospray ionization time-of-flight mass spectrometry as an alternative to two-dimensional gels for the identification and analysis of complex mixtures of intact proteins. *J Chromatogr B Analyt Technol Biomed Life Sci,* 782, 267-89.

LOPEZ, M. F., BERGGREN, K., CHERNOKALSKAYA, E., LAZAREV, A., ROBINSON, M. & PATTON, W. F. (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis,* 21, 3673-83.

MACPHERSON, G. G., WARRELL, M. J., WHITE, N. J., LOOAREESUWAN, S. & WARRELL, D. A. (1985) Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. *Am J Pathol,* 119, 385-401.

MAEGRAITH, B. & FLETCHER, A. (1972) The pathogenesis of mammalian malaria. *Adv Parasitol,* 10, 49-75.

MAROUGA, R., DAVID, S. & HAWKINS, E. (2005) The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal Bioanal Chem,* 382, 669-78.

MARSH, K., FORSTER, D., WARUIRU, C., MWANGI, I., WINSTANLEY, M., MARSH, V., NEWTON, C., WINSTANLEY, P., WARN, P., PESHU, N. & ET AL. (1995) Indicators of life-threatening malaria in African children. *N Engl J Med,* 332, 1399-404.

MATSUI, N. M., SMITH, D. M., CLAUSER, K. R., FICHMANN, J., ANDREWS, L. E., SULLIVAN, C. M., BURLINGAME, A. L. & EPSTEIN, L. B. (1997) Immobilized pH gradient two-dimensional gel electrophoresis and mass spectrometric identification of cytokine-regulated proteins in ME-180 cervical carcinoma cells. *Electrophoresis,* 18, 409-17.

MITULOVIC, G., STINGL, C., SMOLUCH, M., SWART, R., CHERVET, J. P., STEINMACHER, I., GERNER, C. & MECHTLER, K. (2004) Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. *Proteomics,* 4, 2545-57.

MOLLOY, M. P., HERBERT, B. R., WALSH, B. J., TYLER, M. I., TRAINI, M., SANCHEZ, J. C., HOCHSTRASSER, D. F., WILLIAMS, K. L. & GOOLEY, A. A. (1998) Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis,* 19, 837-44.

MOLYNEUX, M. E., TAYLOR, T. E., WIRIMA, J. J. & BORGSTEIN, A. (1989) Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. *Q J Med,* 71, 441-59.

MOTOKAWA, M., KOBAYASHI, H., ISHIZUKA, N., MINAKUCHI, H., NAKANISHI, K., JINNAI, H., HOSOYA, K., IKEGAMI, T. & TANAKA, N. (2002) Monolithic silica columns with various skeleton sizes and through-pore sizes for capillary liquid chromatography. *J Chromatogr A,* 961, 53-63.

NALLY, J. E., WHITELEGGE, J. P., AGUILERA, R., PEREIRA, M. M., BLANCO, D. R. & LOVETT, M. A. (2005) Purification and proteomic analysis of outer membrane vesicles from a clinical isolate of Leptospira interrogans serovar Copenhageni. *Proteomics,* 5, 144-52.

NASH, G. B., O'BRIEN, E., GORDON-SMITH, E. C. & DORMANDY, J. A. (1989) Abnormalities in the mechanical properties of red blood cells caused by Plasmodium falciparum. *Blood,* 74, 855-61.

NEILL, A. L., CHAN-LING, T. & HUNT, N. H. (1993) Comparisons between microvascular changes in cerebral and non-cerebral malaria in mice, using the retinal whole-mount technique. *Parasitology,* 107 (Pt 5), 477-87.

NEILL, A. L. & HUNT, N. H. (1992) Pathology of fatal and resolving Plasmodium berghei cerebral malaria in mice. *Parasitology,* 105 (Pt 2), 165-75.

NEUHOFF, V., AROLD, N., TAUBE, D. & EHRHARDT, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis,* 9, 255-62.

NEWTON, C. R. & KRISHNA, S. (1998) Severe falciparum malaria in children: current understanding of pathophysiology and supportive treatment. *Pharmacol Ther,* 79, 1-53.

NEWTON, C. R. & WARRELL, D. A. (1998) Neurological manifestations of falciparum malaria. *Ann Neurol,* 43, 695-702.

NGUYEN, T. H., DAY, N. P., LY, V. C., WALLER, D., NGUYEN, H. P., BETHELL, D. B., TRAN, T. H. & WHITE, N. J. (1996) Post-malaria neurological syndrome. *Lancet,* 348, 917-21.

NOMURA, F., TOMONAGA, T., SOGAWA, K., OHASHI, T., NEZU, M., SUNAGA, M., KONDO, N., IYO, M., SHIMADA, H. & OCHIAI, T. (2004) Identification of novel and downregulated biomarkers for alcoholism by surface enhanced laser desorption/ionization-mass spectrometry. *Proteomics,* 4, 1187-94.

NUHSE, T. S., STENSBALLE, A., JENSEN, O. N. & PECK, S. C. (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics,* 2, 1234-43.

NUWAYSIR, L. M. & STULTS, J. T. (1993) Electrospray ionization mass spectrometry of phosphopeptides isolated by on-line immobilized metal-ion affinity chromatography. *Journal of the American Society for Mass Spectrometry,* 4, 662.

O'FARRELL, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem,* 250, 4007-21.

OLSEN, J. V., ANDERSEN, J. R., NIELSEN, P. A., NIELSEN, M. L., FIGEYS, D., MANN, M. & WISNIEWSKI, J. R. (2004) HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol Cell Proteomics,* 3, 82-92.

OMENN, G. S., STATES, D. J., ADAMSKI, M., BLACKWELL, T. W., MENON, R., HERMJAKOB, H., APWEILER, R., HAAB, B. B., SIMPSON, R. J., EDDES, J. S., KAPP, E. A., MORITZ, R. L., CHAN, D. W., RAI, A. J., ADMON, A.,

AEBERSOLD, R., ENG, J., HANCOCK, W. S., HEFTA, S. A., MEYER, H., PAIK, Y. K., YOO, J. S., PING, P., POUNDS, J., ADKINS, J., QIAN, X., WANG, R., WASINGER, V., WU, C. Y., ZHAO, X., ZENG, R., ARCHAKOV, A., TSUGITA, A., BEER, I., PANDEY, A., PISANO, M., ANDREWS, P., TAMMEN, H., SPEICHER, D. W. & HANASH, S. M. (2005) Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics,* 5, 3226-45.

PAIN, A., FERGUSON, D. J., KAI, O., URBAN, B. C., LOWE, B., MARSH, K. & ROBERTS, D. J. (2001) Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proc Natl Acad Sci U S A,* 98, 1805-10.

PALMER-TOY, D. E., SARRACINO, D. A., SGROI, D., LEVANGIE, R. & LEOPOLD, P. E. (2000) Direct acquisition of matrix-assisted laser Desorption/Ionization time-of-flight mass spectra from laser capture microdissected tissues. *Clin Chem,* 46, 1513-6.

PAN, S., ZHU, D., QUINN, J. F., PESKIND, E. R., MONTINE, T. J., LIN, B., GOODLETT, D. R., TAYLOR, G., ENG, J. & ZHANG, J. (2007) A combined dataset of human cerebrospinal fluid proteins identified by multi-dimensional chromatography and tandem mass spectrometry. *Proteomics,* 7, 469-73.

PAPADOPOULOS, M. C., ABEL, P. M., AGRANOFF, D., STICH, A., TARELLI, E., BELL, B. A., PLANCHE, T., LOOSEMORE, A., SAADOUN, S., WILKINS, P. & KRISHNA, S. (2004) A novel and accurate diagnostic test for human African trypanosomiasis. *The Lancet,* 363, 1358.

PAPPIN, D. J., HOJRUP, P. & BLEASBY, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol,* 3, 327-32.

PASVOL, G. (2001) Cell-cell interaction in the pathogenesis of severe falciparum malaria. *Clin Med,* 1, 495-500.

PATTON, W. F. (2002) Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci,* 771, 3-31.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis,* 20, 3551-67.

PERROT, M., SAGLIOCCO, F., MINI, T., MONRIBOT, C., SCHNEIDER, U., SHEVCHENKO, A., MANN, M., JENO, P. & BOUCHERIE, H. (1999) Two-dimensional gel protein database of Saccharomyces cerevisiae (update 1999). *Electrophoresis,* 20, 2280-98.

PETRICOIN, E. E., PAWELETZ, C. P. & LIOTTA, L. A. (2002a) Clinical applications of proteomics: proteomic pattern diagnostics. *J Mammary Gland Biol Neoplasia,* 7, 433-40.

PETRICOIN, E. F., 3RD, ORNSTEIN, D. K., PAWELETZ, C. P., ARDEKANI, A., HACKETT, P. S., HITT, B. A., VELASSCO, A., TRUCCO, C., WIEGAND, L., WOOD, K., SIMONE, C. B., LEVINE, P. J., LINEHAN, W. M., EMMERT-BUCK, M. R., STEINBERG, S. M., KOHN, E. C. & LIOTTA, L. A. (2002b) Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst,* 94, 1576-8.

PETRICOIN, E. F., ARDEKANI, A. M., HITT, B. A., LEVINE, P. J., FUSARO, V. A., STEINBERG, S. M., MILLS, G. B., SIMONE, C., FISHMAN, D. A., KOHN, E. C. & LIOTTA, L. A. (2002c) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet,* 359, 572-7.

PETRICOIN, E. F. & LIOTTA, L. A. (2004) SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. *Curr Opin Biotechnol,* 15, 24-30.

PETRICOIN, E. F., ORNSTEIN, D. K., PAWELETZ, C. P., ARDEKANI, A., HACKETT, P. S., HITT, B. A., VELASSCO, A., TRUCCO, C., WIEGAND, L., WOOD, K., SIMONE, C. B., LEVINE, P. J., LINEHAN, W. M., EMMERT-BUCK, M. R., STEINBERG, S. M., KOHN, E. C. & LIOTTA, L. A. (2002d) Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst,* 94, 1576 - 1578.

PETRICOIN, E. F., ZOON, K. C., KOHN, E. C., BARRETT, J. C. & LIOTTA, L. A. (2002e) Clinical proteomics: translating benchside promise into bedside reality. *Nat Rev Drug Discov,* 1, 683-95.

POLANSKI, M. & ANDERSON, N. L. (2006) A list of Candidate Cancer Biomarkers for Targeted Proteomics. *J Biomarker Insights.*

POLDER, T. W., ELING, W. M., CURFS, J. H., JERUSALEM, C. R. & WIJERS-ROUW, M. (1992) Ultrastructural changes in the blood-brain barrier of mice infected with Plasmodium berghei. *Acta Leiden,* 60, 31-46.

PONGPONRATN, E., RIGANTI, M., PUNPOOWONG, B. & AIKAWA, M. (1991) Microvascular sequestration of parasitized erythrocytes in human falciparum malaria: a pathological study. *Am J Trop Med Hyg,* 44, 168-75.

POWELL, M. J. & TIMPERMAN, A. T. (2004) Proteome analysis. *Methods Mol Biol,* 251, 387-400.

PUCHADES, M., HANSSON, S. F., NILSSON, C. L., ANDREASEN, N., BLENNOW, K. & DAVIDSSON, P. (2003) Proteomic studies of potential cerebrospinal fluid protein markers for Alzheimer's disease. *Brain Res Mol Brain Res,* 118, 140-6.

PUTNAM F., W. (1975) *The Plasma Proteins,* New York, Academic Press.

RABILLOUD, T. (1996) Solubilization of proteins for electrophoretic analyses. *Electrophoresis,* 17, 813-29.

RABILLOUD, T. (1998) Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis,* 19, 758-60.

RABILLOUD, T. (1999) Solubilization of proteins in 2-D electrophoresis. An outline. *Methods Mol Biol,* 112, 9-19.

RABILLOUD, T., ADESSI, C., GIRAUDEL, A. & LUNARDI, J. (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis,* 18, 307-16.

RABILLOUD, T., GIANAZZA, E., CATTO, N. & RIGHETTI, P. G. (1990) Amidosulfobetaines, a family of detergents with improved solubilization properties: application for isoelectric focusing under denaturing conditions. *Anal Biochem,* 185, 94-102.

RABILLOUD, T., VALETTE, C. & LAWRENCE, J. J. (1994a) Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis,* 15, 1552-8.

RABILLOUD, T., VUILLARD, L., GILLY, C. & LAWRENCE, J. J. (1994b) Silver-staining of proteins in polyacrylamide gels: a general overview. *Cell Mol Biol (Noisy-le-grand),* 40, 57-75.

RACKER, E., VIOLAND, B., O'NEAL, S., ALFONZO, M. & TELFORD, J. (1979) Reconstitution, a way of biochemical research; some new approaches to membrane-bound enzymes. *Arch Biochem Biophys,* 198, 470-7.

RAES, G., BRYS, L., DAHAL, B. K., BRANDT, J., GROOTEN, J., BROMBACHER, F., VANHAM, G., NOEL, W., BOGAERT, P., BOONEFAES, T., KINDT, A., VAN DEN BERGH, R., LEENEN, P. J., DE BAETSELIER, P. & GHASSABEH, G. H. (2005) Macrophage galactose-type C-type lectins as novel markers for alternatively activated macrophages elicited by parasitic infections and allergic airway inflammation. *J Leukoc Biol,* 77, 321-7.

RAI, A. J., ZHANG, Z., ROSENZWEIG, J., SHIH IE, M., PHAM, T., FUNG, E. T., SOKOLL, L. J. & CHAN, D. W. (2002) Proteomic approaches to tumor marker discovery. *Arch Pathol Lab Med,* 126, 1518-26.

RAMSTROM, M., HAGMAN, C., MITCHELL, J. K., DERRICK, P. J., HAKANSSON, P. & BERGQUIST, J. (2005) Depletion of high-abundant proteins in body fluids prior to liquid chromatography fourier transform ion cyclotron resonance mass spectrometry. *J Proteome Res,* 4, 410-6.

REED, R. C., VERHUEL, A. F., HUNTER, R. L., UDHAYAKUMAR, V., LOUIS-WILEMAN, V., JENNINGS, V. J., JUE, D. L., WOHLHUETER, R. M. & LAL, A. A. (1997) Rapid onset of malaria-induced mortality by immunizations with lipo-peptides: an experimental model to study deleterious immune responses and immunopathology in malaria. *Vaccine,* 15, 65-70.

REISNER, A. H., NEMES, P. & BUCHOLTZ, C. (1975) The use of Coomassie Brilliant Blue G250 perchloric acid solution for staining in electrophoresis and isoelectric focusing on polyacrylamide gels. *Anal Biochem,* 64, 509-16.

REST, J. R. (1982) Cerebral malaria in inbred mice. I. A new model and its pathology. *Trans R Soc Trop Med Hyg,* 76, 410-5.

REYZER, M. L. & CAPRIOLI, R. M. (2005) MALDI mass spectrometry for direct tissue analysis: a new tool for biomarker discovery. *J Proteome Res,* 4, 1138-42.

ROBERTS, D. J., PAIN, A., KAI, O., KORTOK, M. & MARSH, K. (2000) Autoagglutination of malaria-infected red blood cells and malaria severity. *Lancet,* 355, 1427-8.

ROGERS, M. A., CLARKE, P., NOBLE, J., MUNRO, N. P., PAUL, A., SELBY, P. J. & BANKS, R. E. (2003) Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. *Cancer Res,* 63, 6971-83.

ROHLFF, C. (2000) Proteomics in molecular medicine: applications in central nervous systems disorders. *Electrophoresis,* 21, 1227-34.

ROHNER, T. C., STAAB, D. & STOECKLI, M. (2005) MALDI mass spectrometric imaging of biological tissue sections. *Mech Ageing Dev,* 126, 177-85.

ROMEO, M. J., ESPINA, V., LOWENTHAL, M., ESPINA, B. H., PETRICOIN, E. F., 3RD & LIOTTA, L. A. (2005) CSF proteome: a protein repository for potential biomarker identification. *Expert Rev Proteomics,* 2, 57-70.

ROSSI, L., MARTIN, B. M., HORTIN, G. L., WHITE, R. L., FOSTER, M., MOHARRAM, R., STRONCEK, D., WANG, E., MARINCOLA, F. M. & PANELLI, M. C. (2006) Inflammatory protein profile during systemic high dose interleukin-2 administration. *Proteomics,* 6, 709-20.

ROSTY, C., CHRISTA, L., KUZDZAL, S., BALDWIN, W. M., ZAHURAK, M. L., CARNOT, F., CHAN, D. W., CANTO, M., LILLEMOE, K. D., CAMERON, J. L., YEO, C. J., HRUBAN, R. H. & GOGGINS, M. (2002) Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res,* 62, 1868-75.

ROWE, A., OBEIRO, J., NEWBOLD, C. I. & MARSH, K. (1995) Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect Immun,* 63, 2323-6.

ROWE, J. A., MOULDS, J. M., NEWBOLD, C. I. & MILLER, L. H. (1997) P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature,* 388, 292-5.

RUEPP, A., ZOLLNER, A., MAIER, D., ALBERMANN, K., HANI, J., MOKREJS, M., TETKO, I., GULDENER, U., MANNHAUPT, G., MUNSTERKOTTER, M. & MEWES, H. W. (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res,* 32, 5539-45.

SAITO, S. & TSUCHIYA, T. (1984) Characteristics of n-octyl beta-D-thioglucopyranoside, a new non-ionic detergent useful for membrane biochemistry. *Biochem J,* 222, 829-32.

SASAKI, H., YOSHIOKA, N., TAKAGI, Y. & SAKAKI, Y. (1985) Structure of the chromosomal gene for human serum prealbumin. *Gene,* 37, 191-7.

SCHRATTENHOLZ, A. (2004) Proteomics: how to control highly dynamic patterns of millions of molecules and interpret changes correctly? *Drug Discovery Today: Technologies,* 1, 1.

SCHREIBER, G., ALDRED, A. R., JAWOROWSKI, A., NILSSON, C., ACHEN, M. G. & SEGAL, M. B. (1990) Thyroxine transport from blood to brain via transthyretin synthesis in choroid plexus. *Am J Physiol,* 258, R338-45.

SCHREIBER, G., TSYKIN, A., ALDRED, A. R., THOMAS, T., FUNG, W. P., DICKSON, P. W., COLE, T., BIRCH, H., DE JONG, F. A. & MILLAND, J. (1989) The acute phase response in the rodent. *Ann N Y Acad Sci,* 557, 61-85; discussion 85-6.

SCUDERI, P., DORR, R. T., LIDDIL, J. D., FINLEY, P. R., MELTZER, P., RAITANO, A. B. & RYBSKI, J. (1989) Alpha-globulins suppress human leukocyte tumor necrosis factor secretion. *Eur J Immunol,* 19, 939-42.

SEEHUSEN, D. A., REEVES, M. M. & FOMIN, D. A. (2003) Cerebrospinal fluid analysis. *Am Fam Physician,* 68, 1103-8.

SEKIYA, S., WADA, Y. & TANAKA, K. (2005) Derivatization for stabilizing sialic acids in MALDI-MS. *Anal Chem,* 77, 4962-8.

SENANAYAKE, N. & DE SILVA, H. J. (1994) Delayed cerebellar ataxia complicating falciparum malaria: a clinical study of 74 patients. *J Neurol,* 241, 456-9.

SHAW, M. M. & RIEDERER, B. M. (2003) Sample preparation for two-dimensional gel electrophoresis. *Proteomics,* 3, 1408-17.

SHEN, Z., WANT, E. J., CHEN, W., KEATING, W., NUSSBAUMER, W., MOORE, R., GENTLE, T. M. & SIUZDAK, G. (2006) Sepsis plasma protein profiling with

immunodepletion, three-dimensional liquid chromatography tandem mass spectrometry, and spectrum counting. *J Proteome Res,* 5, 3154-60.

SHEVCHENKO, A., JENSEN, O. N., PODTELEJNIKOV, A. V., SAGLIOCCO, F., WILM, M., VORM, O., MORTENSEN, P., SHEVCHENKO, A., BOUCHERIE, H. & MANN, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci U S A,* 93, 14440-5.

SHI, Y., XIANG, R., HORVATH, C. & WILKINS, J. A. (2004) The role of liquid chromatography in proteomics. *J Chromatogr A,* 1053, 27-36.

SINZ, A., BANTSCHEFF, M., MIKKAT, S., RINGEL, B., DRYNDA, S., KEKOW, J., THIESEN, H. J. & GLOCKER, M. O. (2002) Mass spectrometric proteome analyses of synovial fluids and plasmas from patients suffering from rheumatoid arthritis and comparison to reactive arthritis or osteoarthritis. *Electrophoresis,* 23, 3445-56.

SIUZDAK, G. (1996) *Mass Spectrometry for Biotechnology,* San Diego, Academic Press

SNOW, R. W., GUERRA, C. A., NOOR, A. M., MYINT, H. Y. & HAY, S. I. (2005) The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature,* 434, 214-7.

SORACE, J. M. & ZHAN, M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics,* 4, 24.

SOTTANI, C., FIORENTINO, M. & MINOIA, C. (1997) Matrix performance in matrix-assisted laser desorption/ionization for molecular weight determination in sialyl and non-sialyl oligosaccharide proteins. *Rapid Commun Mass Spectrom,* 11, 907-13.

SOUTHWELL, B. R., DUAN, W., ALCORN, D., BRACK, C., RICHARDSON, S. J., KOHRLE, J. & SCHREIBER, G. (1993) Thyroxine transport to the brain: role of protein synthesis by the choroid plexus. *Endocrinology,* 133, 2116-26.

SPECTOR, T. (1978) Refinement of the coomassie blue method of protein quantitation. A simple and linear spectrophotometric assay for less than or equal to 0.5 to 50 microgram of protein. *Anal Biochem,* 86, 142-6.

STEEL, L. F., SHUMPERT, D., TROTTER, M., SEEHOLZER, S. H., EVANS, A. A., LONDON, W. T., DWEK, R. & BLOCK, T. M. (2003) A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics,* 3, 601-9.

STERNER, J. L., JOHNSTON, M. V., NICOL, G. R. & RIDGE, D. P. (2000) Signal suppression in electrospray ionization Fourier transform mass spectrometry of multi-component samples. *J Mass Spectrom,* 35, 385-91.

STOECKLI, M., FARMER, T. B. & CAPRIOLI, R. M. (1999) Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument. *J Am Soc Mass Spectrom,* 10, 67-71.

TABONE, M. D., MUANZA, K., LYAGOUBI, M., JARDEL, C., PIED, S., AMEDEE-MANESME, O., GRAU, G. E. & MAZIER, D. (1992) The role of interleukin-6 in vitamin A deficiency during Plasmodium falciparum malaria and possible consequences for vitamin A supplementation. *Immunology,* 75, 553-4.

TAM, S. W., PIRRO, J. & HINERFELD, D. (2004) Depletion and fractionation technologies in plasma proteomic analysis. *Expert Rev Proteomics,* 1, 411-20.

TAMMEN, H., SCHULTE, I., HESS, R., MENZEL, C., KELLMANN, M. & SCHULZ-KNAPPE, P. (2005) Prerequisites for peptidomic analysis of blood samples: I. Evaluation of blood specimen qualities and determination of technical performance characteristics. *Comb Chem High Throughput Screen,* 8, 725-33.

TANAKA, N., KOBAYASHI, H., NAKANISHI, K., MINAKUCHI, H. & ISHIZUKA, N. (2001) Monolithic LC columns. *Anal Chem,* 73, 420A-429A.

TERRY, D. E., UMSTOT, E. & DESIDERIO, D. M. (2004) Optimized sample-processing time and peptide recovery for the mass spectrometric analysis of protein digests. *J Am Soc Mass Spectrom,* 15, 784-94.

THONGBOONKERD, V., KLEIN, E. & KLEIN, J. B. (2004) Sample preparation for 2-D proteomic analysis. *Contrib Nephrol,* 141, 11-24.

THUMWOOD, C. M., HUNT, N. H., CLARK, I. A. & COWDEN, W. B. (1988) Breakdown of the blood-brain barrier in murine cerebral malaria. *Parasitology,* 96 (Pt 3), 579-89.

THUMWOOD, C. M., HUNT, N. H., COWDEN, W. B. & CLARK, I. A. (1989) Antioxidants can prevent cerebral malaria in Plasmodium berghei-infected mice. *Br J Exp Pathol,* 70, 293-303.

THURNHAM, D. I. & SINGKAMANI, R. (1991) The acute phase response and vitamin A status in malaria. *Trans R Soc Trop Med Hyg,* 85, 194-9.

TONGE, R., SHAW, J., MIDDLETON, B., ROWLINSON, R., RAYNER, S., YOUNG, J., POGNAN, F., HAWKINS, E., CURRIE, I. & DAVISON, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics,* 1, 377-96.

TRESELER, K. (1995) *Clinical Laboratory Diagnostic Tests*
Connecticut, USA, Appleton and Lange.

TREUTIGER, C. J., HEDLUND, I., HELMBY, H., CARLSON, J., JEPSON, A., TWUMASI, P., KWIATKOWSKI, D., GREENWOOD, B. M. & WAHLGREN, M. (1992) Rosette formation in Plasmodium falciparum isolates and anti-rosette activity of sera from Gambians with cerebral or uncomplicated malaria. *Am J Trop Med Hyg,* 46, 503-10.

TURNER, G. (1997) Cerebral malaria. *Brain Pathol,* 7, 569-82.

TURNER, G. D., MORRISON, H., JONES, M., DAVIS, T. M., LOOAREESUWAN, S., BULEY, I. D., GATTER, K. C., NEWBOLD, C. I., PUKRITAYAKAMEE, S., NAGACHINTA, B. & ET AL. (1994) An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J Pathol,* 145, 1057-69.

TWYMAN, R. M. (2004) *Principles of Proteomics*, BIOS Scientific Publishers.

TYERS, M. & MANN, M. (2003) From genomics to proteomics. *Nature,* 422, 193-7.

UNLU, M., MORGAN, M. E. & MINDEN, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis,* 18, 2071-7.

VALERIO, A., BASSO, D., MAZZA, S., BALDO, G., TIENGO, A., PEDRAZZOLI, S., SERAGLIA, R. & PLEBANI, M. (2001) Serum protein profiles of patients with pancreatic cancer and chronic pancreatitis: searching for a diagnostic protein pattern. *Rapid Commun Mass Spectrom,* 15, 2420-5.

WADSWORTH, J. T., SOMERS, K. D., CAZARES, L. H., MALIK, G., ADAM, B. L., STACK, B. C., JR., WRIGHT, G. L., JR. & SEMMES, O. J. (2004) Serum protein profiles to identify head and neck cancer. *Clin Cancer Res,* 10, 1625-32.

WALCHER, W., OBERACHER, H., TROIANI, S., HOLZL, G., OEFNER, P., ZOLLA, L. & HUBER, C. G. (2002) Monolithic capillary columns for liquid chromatography-electrospray ionization mass spectrometry in proteomic and genomic research. *J Chromatogr B Analyt Technol Biomed Life Sci,* 782, 111-25.

WALDUCK, A., RUDEL, T. & MEYER, T. F. (2004) Proteomic and gene profiling approaches to study host responses to bacterial infection. *Curr Opin Microbiol,* 7, 33-8.

WANG, Y. Y., CHENG, P. & CHAN, D. W. (2003) A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis. *Proteomics,* 3, 243-8.

WARRELL, D. A. (1987) Pathophysiology of severe falciparum malaria in man. *Parasitology,* 94 Suppl, S53-76.

WARUIRU, C. M., NEWTON, C. R. J. C., FORSTER, D., NEW, L., WINSTANLEY, P., MWANGI, I., MARSH, V., WINSTANLEY, M., SNOW, R. W. & MARSH, K. (1996) Epileptic seizures and malaria in Kenyan children. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 90, 152-155.

WASHBURN, M. P., WOLTERS, D. & YATES, J. R., 3RD (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol,* 19, 242-7.

WATTS, J. D., AFFOLTER, M., KREBS, D. L., WANGE, R. L., SAMELSON, L. E. & AEBERSOLD, R. (1994) Identification by electrospray ionization mass spectrometry of the sites of tyrosine phosphorylation induced in activated Jurkat T cells on the protein tyrosine kinase ZAP-70. *J Biol Chem,* 269, 29520-9.

WEINBERGER, S. R., BOSCHETTI, E., SANTAMBIEN, P. & BRENAC, V. (2002) Surface-enhanced laser desorption-ionization retentate chromatography mass spectrometry (SELDI-RC-MS): a new method for rapid development of process chromatography conditions. *J Chromatogr B Analyt Technol Biomed Life Sci,* 782, 307-16.

WHO (2003) Shaping the Future. *World Health Report.* Geneva, World Health Organisation.

WILKINS, M. R., APPEL, R. D., VAN EYK, J. E., CHUNG, M. C., GORG, A., HECKER, M., HUBER, L. A., LANGEN, H., LINK, A. J., PAIK, Y. K., PATTERSON, S. D., PENNINGTON, S. R., RABILLOUD, T., SIMPSON, R. J., WEISS, W. & DUNN, M. J. (2006) Guidelines for the next 10 years of proteomics. *Proteomics,* 6, 4-8.

WILM, M., SHEVCHENKO, A., HOUTHAEVE, T., BREIT, S., SCHWEIGERER, L., FOTSIS, T. & MANN, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature,* 379, 466-9.

WILSON, D. L., HALL, M. E., STONE, G. C. & RUBIN, R. W. (1977) Some improvements in two-dimensional gel electrophoresis of proteins. Protein mapping of eukaryotic tissue extracts. *Anal Biochem,* 83, 33-44.

WOLTERS, D. A., WASHBURN, M. P. & YATES, J. R., 3RD (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem,* 73, 5683-90.

WON, Y., SONG, H. J., KANG, T. W., KIM, J. J., HAN, B. D. & LEE, S. W. (2003) Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons. *Proteomics,* 3, 2310-6.

WU, C. C. & MACCOSS, M. J. (2002) Shotgun proteomics: tools for the analysis of complex biological systems. *Curr Opin Mol Ther,* 4, 242-50.

WU, F. S. & WANG, M. Y. (1984) Extraction of proteins for sodium dodecyl sulfate-polyacrylamide gel electrophoresis from protease-rich plant tissues. *Anal Biochem,* 139, 100-3.

XU, B. J., CAPRIOLI, R. M., SANDERS, M. E. & JENSEN, R. A. (2002) Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J Am Soc Mass Spectrom,* 13, 1292-7.

YATES, J. R., 3RD (1998) Database searching using mass spectrometry data. *Electrophoresis,* 19, 893-900.

YATES, J. R., 3RD, CARMACK, E., HAYS, L., LINK, A. J. & ENG, J. K. (1999) Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. *Methods Mol Biol,* 112, 553-69.

YATES, J. R., 3RD, ENG, J. K. & MCCORMACK, A. L. (1995a) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem,* 67, 3202-10.

YATES, J. R., 3RD, ENG, J. K., MCCORMACK, A. L. & SCHIELTZ, D. (1995b) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem,* 67, 1426-36.

YE, B., CRAMER, D. W., SKATES, S. J., GYGI, S. P., PRATOMO, V., FU, L., HORICK, N. K., LICKLIDER, L. J., SCHORGE, J. O., BERKOWITZ, R. S. & MOK, S. C. (2003) Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin Cancer Res,* 9, 2904-11.

YUAN, X. & DESIDERIO, D. M. (2005a) Proteomics analysis of human cerebrospinal fluid. *Journal of Chromatography B,* 815, 179.

YUAN, X. & DESIDERIO, D. M. (2005b) Proteomics analysis of human cerebrospinal fluid. *J Chromatogr B Analyt Technol Biomed Life Sci,* 815, 179-89.

YUAN, X., RUSSELL, T., WOOD, G. & DESIDERIO, D. M. (2002) Analysis of the human lumbar cerebrospinal fluid proteome. *Electrophoresis,* 23, 1185-96.

ZHANG, J., GOODLETT, D. R., PESKIND, E. R., QUINN, J. F., ZHOU, Y., WANG, Q., PAN, C., YI, E., ENG, J., AEBERSOLD, R. H. & MONTINE, T. J. (2005a) Quantitative proteomic analysis of age-related changes in human cerebrospinal fluid. *Neurobiol Aging,* 26, 207-27.

ZHANG, J., GOODLETT, D. R., QUINN, J. F., PESKIND, E., KAYE, J. A., ZHOU, Y., PAN, C., YI, E., ENG, J., WANG, Q., AEBERSOLD, R. H. & MONTINE, T. J. (2005b) Quantitative proteomics of cerebrospinal fluid from patients with Alzheimer disease. *J Alzheimers Dis,* 7, 125-33; discussion 173-80.

ZOLG, J. W. & LANGEN, H. (2004) How industry is approaching the search for new diagnostic markers and biomarkers. *Mol Cell Proteomics,* 3, 345-54.

ZOLOTARJOVA, N., MARTOSELLA, J., NICOL, G., BAILEY, J., BOYES, B. E. & BARRETT, W. C. (2005) Differences among techniques for high-abundant protein depletion. *Proteomics,* 5, 3304-13.

APPENDIX I List of human proteins identified in plasma

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
LPHUB apolipoprotein B-100 precursor - human -induced protein 2	71789					4	50.00
A25518 glycogen phosphorylase (EC 2.4.1.1 , hepatic - human	87575			4	12.73	4	12.73
B34611 3',5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35 alpha chain - human	105117	3	5.75	2	4.37	3	5.75
A34653 cell adhesion protein SQM1 - human	105595					9	16.73
A40267 interleukin-5 receptor alpha chain precursor - human	106824	8	24.73	2	6.83	16	50.00
APA4_HUMAN Apolipoprotein A-IV precursor (Apo-AIV eta pr	114006					2	50.00
G6PD_HUMAN Glucose-6-phosphate 1-dehydrogenase (G6PD	120731					9	25.57
PGH1_HUMAN Prostaglandin G/H synthase 1 precursor (Cyclooxygenase - 1 (COX-1 (	129899	5	8.54	14	14.43	10	12.21
SPCB_HUMAN Spectrin beta chain, erythrocyte (Beta-I spectrin	134798					2	7.73
sperm activating protein subunit I, apolipoprotein A1, SPAP subunit I [hum	235865			3	50.00		
G42075 finger protein (clone ZnFP34 - human (fragment id And	321128			2	6.54	4	18.73
S38532 protein 14-3-3 eta chain - human	481365					4	7.28

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
S41768 splicing factor homolog - human	543010					7	18.73
JC2401 PMS6 homolog mismatch repair protein - human	1082699			3	7.73		
CENF_HUMAN CENP-F kinetochore protein (Centromere protein F (Mitosin (AH ant	1345731					2	6.82
K2C1_HUMAN Keratin, type II cytoskeletal 1 (Cytokeratin 1 (K1 (CK 1 (67 kDa	1346343	5	28.77			4	9.94
A57099 DNA-activated protein kinase, catalytic subunit - human	1362789					2	8.74
A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor (Alpha-1	1703025	22	44.24	28	48.73	23	45.21
ODO2_HUMAN Dihydrolipoamide succinyltransferase component of 2-oxoglutarate de	1709442					5	20.25
I38344 titin, cardiac muscle [validated] - human iciency viru	2136280					12	18.73
S69339 Ig heavy chain V region precursor - human ain-binding	2146957	25	14.44	51	25.58	46	40.68
DN2L_HUMAN DNA2-like homolog (DNA replication helicase-like homolog	2506893					8	12.72
Cdc42 From Human, Nmr, 20 Structures n (fragment	2624582			2	6.73		
A Chain A, Crystal Structure Of Human Apolipoprotein A-I	2914175			20	70.73		
PR16_HUMAN Pre-mRNA splicing factor ATP-dependent RNA helicase PRP16 (ATP-depe	3024898					3	12.63

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
CTOG_HUMAN CH-TOG protein (Colonic and hepatic tumor ove	3121951					3	8.74
Y008_HUMAN Hypothetical protein KIAA0008	3183208					2	7.43
DPOZ_HUMAN DNA polymerase zeta catalytic subunit (hREV3	3913527					2	8.24
H Chain H, Comparison Of The Three-Dimensional Structures Of A Humanized And A	4139696	23	18.23	76	30.24	5	10.28
C Chain C, Human Serum Transferrin, Recombinant N-Terminal Lobe, Apo Form	4389232	5	15.95	6	20.18	9	30.24
Human Serum Albumin In A Complex With Myristic Acid And Tri-Iodobenzoic Acid	4389275	89	38.35	425	72.23	188	60.27
albumin precursor; PRO0883 protein [Homo sapiens]	4502027	6	12.24	3	6.84	17	15.93
membrane alanine aminopeptidase precursor; microsomal aminopeptidase; A	4502095					3	7.83
brain-specific angiogenesis inhibitor 2; Brain-sp	4502357					11	15.73
calcium/calmodulin-dependent protein kinase IV; brain Ca(2	4502557					3	12.24
desmocollin 3, isoform Dsc3a preproprotein; desmocollin 4 [Homo sapiens	4503399					3	8.72
early endosome antigen 1, 162kD; early endosome-associated protein [Hom	4503469					2	6.93
eukaryotic translation initiation factor 4E; eukaryotic translation ini	4503535					2	6.73

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
coagulation factor VIII, procoagulant component, isoform a precursor;	4503647					6	20.25
solute carrier family 27 (fatty acid transporter , member 2; very long-	4503653					3	12.37
filamin 1 (actin-binding protein-280 ; filamin A, alpha (actin-binding	4503745					2	10.28
phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide	4503915					2	10.24
GATA binding protein 4; GATA-binding protein 4 [Homo sapiens]	4503931					3	7.74
glutamate receptor, metabotropic 2 precursor [Homo sapiens]	4504137					2	6.65
guanylate cyclase 2F; RetGC-2; guanylate cyclase 2D-like, membrane (ret	4504219	123	15.94			24	10.84
delta globin [Homo sapiens]	4504351			4	24.35		
insulin-like growth factor 2 receptor; Insulin-like growth factor-2 rec	4504611					2	30.70
gi|4504631|ref|NP_003630.1|	4504631					2	8.74
insulin receptor substrate 2 [Homo sapiens]	4504731			5	4.89	13	10.24
inositol 1,4,5-triphosphate receptor, type 2 [Homo sapiens]	4504793					2	6.74
Down syndrome critical region protein 2; chromosome 21 leucine-rich pro	4505023					3	8.74
receptor interacting protein 140; Nuclear recepto	4505455					2	12.67
orosomucoid 2; alpha-1-acid glycoprotein, type 2 [Homo sapiens]	4505529			13	20.29		
plectin 1, intermediate filament binding protein	4505877					4	15.59

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
proteasome alpha 2 subunit; proteasome subunit HC3; proteasome componen	4506181			3	12.39		
protein tyrosine phosphatase, receptor type, F, i	4506311	4	12.27			8	18.63
ribosomal protein L24; 60S ribosomal protein L24; ribosomal protein L30	4506619					2	7.24
ribosomal protein L32; 60S ribosomal protein L32	4506635					7	9.24
restin (Reed-Steinberg cell-expressed intermediate filament-associated	4506751					2	7.83
secretogranin II precursor; Chromogranin C (secretogranin II ; secreton	4506801					24	15.24
sodium channel, nonvoltage-gated 1 alpha [Homo sapiens]	4506815					2	8.4
sorbitol dehydrogenase [Homo sapiens]	4507155					3	12.22
squalene monooxygenase [Homo sapiens]	4507197					2	8.74
treacle; Treacher Collins-Franceschetti syndrome	4507411					2	6.49
thyrotrophic embryonic factor; Thyrotroph embryonic factor [Homo sapien	4507431					2	7.12
transforming growth factor, beta-induced, 68kDa; corneal dystrophy; ker	4507467					2	8.93
translocated promoter region (to activated MET on	4507659					2	12.47
zinc finger protein 84 (HPF2 [Homo sapiens]	4508037					2	8.63
zinc finger protein 91 (HPF7, HTF10 [Homo sapiens]	4508041			5	10.44	2	7.81

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
guanine nucleotide exchange factor p532 [Homo sapiens]	4557026					3	7.55
alpha 2 macroglobulin precursor [Homo sapiens]	4557225	2	5.61	17	12.24	6	6.72
apolipoprotein A-I precursor [Homo sapiens]	4557321			8	8.21		
cell division cycle 20; cell division cycle 20, S.cerevisiae homolog [H	4557437					2	6.27
chromodomain helicase DNA binding protein 4; Mi-2b [Homo sapiens]	4557453					3	8.63
cylicin 2 [Homo sapiens]	4557509					2	5.63
fatty acid binding protein 5 (psoriasis-associated ; E-FABP [Homo sapie	4557581					2	5.83
alpha-L-iduronidase precursor [Homo sapiens]	4557661					2	5.24
Propionyl-Coenzyme A carboxylase, alpha polypeptide precursor [Homo sap	4557833	8	10.23	8	10.23		
transferrin [Homo sapiens]	4557871	2	4.82	11	12.24	6	7.63
HLA-B associated transcript-1; DEAD-box protein; nuclear RNA helicase (	4758112					2	6.47
desmoplakin (DPI, DPII ; Desmoplakin [Homo sapiens]	4758200					7	8.27
forkhead box G1A; forkhead (Drosophila -like 2; Forkhead, drosophila, h	4758390	2	6.67			2	6.67
golgi autoantigen, golgin subfamily b, macrogolgin (with transmembrane	4758454					3	5.83
ankyrin repeat-containing protein Krit1 [Homo sapiens]	4758658					3	8.71
myosin IXB [Homo sapiens]	4758750					5	6.54
NADH dehydrogenase (ubiquinone Fe-S protein 3, 30kDa (NADH-coenzyme Q	4758788					3	6.45

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
phosphoinositide-3-kinase, class 2, gamma polypep	4758924					4	7.41
RAB28, member RAS oncogene family [Homo sapiens]	4758994					3	6.73
retinol-binding protein 2, cellular [Homo sapiens]	4759028			2	6.24	3	8.73
SRY (sex determining region Y -box 14; SRY (sex-determining region Y -b	4759162	2	2.23			4	6.94
VAMP (vesicle-associated membrane protein -associated protein B and C;	4759302					2	8.73
glutamine-fructose-6-phosphate transaminase 2; glutamine: fructose-6-ph	4826742					2	9.63
matrix metalloproteinase 14 preproprotein; membrane-type matrix metallo	4826834					2	9.44
solute carrier family 12 (potassium/chloride transporters , member 4 [H	4827006					3	7.24
ubiquitin specific protease 8 [Homo sapiens]	4827054					4	12.25
chondroitin sulfate proteoglycan 6 (bamacan ; human chromosome-associat	4885399					5	15.61
huntingtin interacting protein 2; ubiquitin-conjugating enzyme E2-25 KD	4885417					3	19.27
nucleolin [Homo sapiens]	4885511					2	18.73
wingless-type MMTV integration site family, member 1 precursor; WNT-1 p	4885655					4	20.24
microtubule-associated protein 1B, isoform 1 [Hom	5174525					6	10.04

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
mitochondrial intermediate peptidase [Homo sapiens]	5174567					2	6.87
nuclear mitotic apparatus protein 1 [Homo sapiens	5453820	2	6.72			2	6.72
ubiquitin specific protease 16; human ubiquitin processing protease [Ho	5454156					8	10.24
cysteinyl leukotriene receptor 1 [Homo sapiens]	5729798					2	10.5
v-rel reticuloendotheliosis viral oncogene homolog B, nuclear factor of	5730007	88	15.87			20	10.32
AFG3 ATPase family gene 3-like 2; ATPase family gene 3-like 2; ATPase f	5802970					2	6.75
zinc finger protein 220; Monocytic leukemia zinc finger protein [Homo s	5803098					4	12.23
protein kinase CHK2 isoform a; checkpoint-like protein CHK2; serine/thr	6005850					2	9.57
human leucocyte antigen A [Homo sapiens] ursor; co	6138770					4	16.47
CTRO_HUMAN Citron protein (Rho-interacting, serine/threonine kinase 21	6225217					4	11.24
Protein required for cell viability; Ygl245wp [Saccharomyces cerevisiae	6321192					4	12.35
component of oligomeric golgi complex 2; brefeldin A-sensitive, periphe	6678676					5	20.24
golgi autoantigen, golgin subfamily a, 4; golgin	6715600	2	6.63			7	15.75
bromodomain adjacent to zinc finger domain, 1B; Williams-Beuren syndrom	6755993					8	18.25
low density lipoprotein-related protein 2; megali	6806919					9	20.25

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
CASP8 associated protein 2; FLICE associated huge; human FLASH; FLASH h	6912288	5	8.32	2	6.24	10	15.73
zinc finger protein 281; ZNP-99 transcription factor [Homo sapiens]	6912752			2	5.92		
fibronectin leucine rich transmembrane protein 2 [Homo sapiens]	7019381					11	12.57
A Chain A, Human Bleomycin Hydrolase.	7245509					12	20.93
S64732 scaffold attachment factor B - human human (fragment	7430315					13	16.00
T12687 ALR protein homolog - fruit fly (Drosophila melanogaster	7511805					14	12.87
I38346 elastic titin - human (fragment	7512404					15	6.24
JC5276 HXC-26 protein - human	7512493					16	7.83
T17289 hypothetical protein DKFZp434P1650.1 - human (fragment	7512694	5	6.93			17	15.87
T12462 hypothetical protein DKFZp564I122.1 - human (fragment	7512752	6	8.72			18	12.87
T12457 hypothetical protein DKFZp564O0423.1 - human (fragment	7512780	8	10.23			19	13.45
T08738 hypothetical protein DKFZp586E0518.1 - human (fragment	7512872	6	7.24			20	18.25
T01357 hypothetical protein H_GS368F15.1 - human (fragment	7512961					21	17.23
T00095 hypothetical protein KIAA0470 - human	7512995					22	16.24
T00326 hypothetical protein KIAA0550 - human	7513013			5	7.32	23	20.24
T00338 hypothetical protein KIAA0570 - human	7513025	2	6.21				

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
T00343 hypothetical protein KIAA0584 - human (fragment	7513030					2	7.64
JE0261 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.- - human	7513171					2	6.48
JC7118 headpin serine proteinase inhibitor - human	7522623					2	5.71
gi|7657122|ref|NP_056536.1|	7657122					3	5.67
group III secreted phospholipase A2 [Homo sapiens]	7657126					3	5.73
DnaJ (Hsp40 homolog, subfamily C, member 8; splicing factor similar to	7657611	3	5.72			8	7.73
centaurin beta1; KIAA0050 gene product; Arf GAP w	7661880					2	6.17
KIAA0053 gene product [Homo sapiens]	7661882					8	10.71
KIAA0233 gene product [Homo sapiens]	7662014			2	6.71		
TBC1 domain family, member 4; KIAA0603 gene product; TBC (Tre-2, BUB2,	7662198					9	10.17
GTPase regulator associated with the focal adhesion kinase pp125 [Homo	7662208					10	12.27
KIAA0672 gene product [Homo sapiens]	7662242					11	13.71
epithelial protein lost in neoplasm beta; sterol regulatory element bin	7705373			2	11.21	12	12.21
beta-ureidopropionase [Homo sapiens]	7706509					13	14.21
CU05_HUMAN Protein C21orf5 n 11 homolog A isoform 1; dou	8134355					14	13.23
tuftelin interacting protein 11; similar to tuftelin-interacting protei	8393259					15	14.87

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
guanylate cyclase 1, soluble, beta 2 [Homo sapiens]	8393507					16	18.63
M3K1_HUMAN Mitogen-activated protein kinase kinase kinas	8488988					17	16.74
hypothetical protein DKFZp761P1010 [Homo sapiens]	8922179					18	17.23
F-box only protein 6; F-box protein Fbx6; F-box protein FBG2 [Homo sapi	8922188					19	18.61
TRIO_HUMAN Triple functional domain protein (PTPRF inter	8928460					20	13.43
orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1 ; a	9257232	5	12.24				
PR domain containing 11; PR-domain containing protein 11 [Homo sapiens]	9910506					3	16.47
ATP-binding cassette, sub-family B, member 10 [Homo sapiens]	9961244	5	8.64	10	12.34	9	11.27
U2 small nuclear ribonucleoprotein auxiliary fac	10835077					2	10.53
squamous cell carcinoma antigen recognized by T cells 1; U4/U6.U5 tri-	10863889					2	10.24
neurogenic differentiation 4 [Homo sapiens]	10863999					3	8.71
insulin-like growth factor 1 (somatomedin C ; in	11024682					7	18.73
N-ethylmaleimide-sensitive factor [Homo sapiens]	11079228					4	16.43
T46352 hypothetical protein DKFZp434C0816.1 - human (fragment	11281986					4	15.24
T43445 hypothetical protein DKFZp434A2017.1 - human (fragmen	11359896					4	13.31
T42701 hypothetical protein DKFZp434G156.1 - human (fragment	11360016					2	10.20

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
T46337 hypothetical protein DKFZp434O2413.1 - human (fragment	11360154					3	9.72
T43481 probable mucin DKFZp434C196.1 - human (fragment	11360318			4	12.24		
hepatocellular carcinoma antigen gene 520 [Homo	11545811					2	11.21
fibrinogen, alpha chain, isoform alpha preproprotein; fibrinogen, A al	11761629					4	9.81
JMJ_HUMAN Jumonji protein	12230209					2	6.71
islet cell autoantigen 1 isoform 1; islet cell autoantigen 1 (69kD ; i	12545395					2	5.67
ZN93_HUMAN Zinc finger protein 93 (Zinc finger protein HTF34	12643428			3	8.72		
Z236_HUMAN Zinc finger protein 236	12643896					4	6.43
AF6_HUMAN AF-6 protein	12644018					2	5.61
P2157	12644088					3	7.21
ATC1_HUMAN Calcium-transporting ATPase type 2C, member 1 (ATPase 2C1 (ATP-de	12644373					2	7.32
ADP-ribosyltransferase 4 [Homo sapiens]	12711662					2	8.23
glucosamine-6-phosphate isomerase; glucosamine-6-phosphate deaminase;	13027378					3	6.83
ABC8_HUMAN ATP-binding cassette, sub-family B, member 8, mitochondrial precur	13123950					2	5.75
smooth muscle myosin heavy chain 11, isoform SM1 [Homo sapiens]	13124879					4	12.34
zinc finger protein 117 (HPF9 ; Zinc finger protein-117 [Homo sapiens]	13374557			2	10.14		
CA2B_HUMAN Collagen alpha 2(XI chain precursor rosophi	13432104					7	11.28

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
wingless-type MMTV integration site family, member 5B precursor; WNT-5	14249180					2	6.53
B Chain B, Crystal Structure Of A Human Fcg Receptor In Complex With An Fc Fra	14277820	9	12.47				
ATB4_HUMAN Plasma membrane calcium-transporting ATPase 4 (PMCA4 (Plasma memb	14286105					4	8.76
general transcription factor II, i, isoform 4; B	14670356					8	7.84
Y226_HUMAN Hypothetical protein KIAA0226	14917097					2	6.57
VAC1_HUMAN Vasopressin-activated calcium-mobilizing receptor (VACM-1 (Cullin	14917099					2	10.81
CNO7_HUMAN CCR4-NOT transcription complex, subunit 7 (CCR4-associated factor	15213950					2	17.24
zinc finger, imprinted 3 [Homo sapiens]	16418391					4	12.43
dynein, axonemal, heavy polypeptide 11; dynein,	16507235					2	6.41
myosin light chain kinase isoform 1; myosin ligh	16950611					2	10.32
KLF8_HUMAN Krueppel-like factor 8 (Zinc finger protein 741 (Basic kruppel-li	17366251					2	11.24
SPCQ_HUMAN Spectrin beta chain, brain 3 (Spectrin, non-	17368942	152	24.32	13	13.47	38	20.47
RM56_HUMAN Mitochondrial 39S ribosomal protein L56 (MRP-L56 (Serine beta lac	17380287					2	11.17

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
MOT3_HUMAN Monocarboxylate transporter 3 (MCT 3	17433295					3	14.17
vacuolar protein sorting 33B (yeast homolog [Homo sapiens]	18105058					2	11.24
L Chain L, Crystal Structure Of Tissue Factor In Complex With Humanized Fab D3	18655500	21	13.75	33	22.57	8	10.64
enhancer of zeste homolog 1 (Drosophila ; enhancer of zeste (Drosophil	19923202			2	10.73		
AT-binding transcription factor 1; AT motif-binding factor 1 [Homo sap	19923287					2	11.24
oxysterol-binding protein-like protein 9 isoform	20070331					2	10.84
adaptor-related protein complex 3, delta 1 subunit; adaptin, delta [Ho	20127438					2	9.86
DMN_HUMAN Desmuslin	20137613					2	5.47
PRES_HUMAN Prestin	20139418					3	7.57
MCAK_HUMAN Mitotic centromere-associated kinesin (MCAK	20141607					2	6.67
NDR2_HUMAN NDRG2 protein (Syld709613 protein	20141615	10	11.74	5	10.87	16	24.47
ribosomal protein S6 kinase, 90kDa, polypeptide 1; ribosomal protein S	20149547			7	15.67		
HT2A_HUMAN Zinc-finger protein HT2A (72 kDa Tat-interacting protein (Tripart	20178303					4	6.73
NCK2_HUMAN Cytoplasmic protein NCK2 (NCK adaptor protei	20532395					2	5.73
similar to Inhibitor of nuclear factor kappa B k	20538863					2	7.21

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
similar to hypothetical protein DKFZp762B245.1 -	20550654					3	5.74
B Chain B, Human Tbx3, A Transcription Factor Responsible For Ulnar-Mammary Sy	20663774					4	7.83
PYRIN-containing APAF1-like protein 5 [Homo sapi	21264320					3	6.54
translocase of outer mitochondrial membrane 34; outer mitochondrial me	21361356					2	7.57
TM31_HUMAN Tripartite motif protein 31	21363032					2	8.21
ANK3_HUMAN Ankyrin 3 (ANK-3 (Ankyrin G	21759000					3	6.74
TSP-EAR [Homo sapiens]	22001420					4	9.21
CAA64752.1	22261813					2	6.51
BA2B_HUMAN Bromodomain adjacent to zinc finger domain 2B (hWALp4	22653668					12	18.57
neurexophilin 1 [Homo sapiens]	23097338					2	7.81
sorting nexin 1 isoform a; sorting nexin 1A [Homo sapiens]	23111034					5	9.61
CHD5_HUMAN Chromodomain-helicase-DNA-binding protein 5 (CHD-5	23396493					3	5.24
RRB1_HUMAN Ribosome binding protein 1 (Ribosome receptor protein (180 kDa ri	23822112					8	18.71
HIP1_HUMAN Huntingtin interacting protein 1 (HIP-I	23831094			8	6.41	8	6.41
SEC6_HUMAN Exocyst complex component Sec6 ÿ	24418663					2	8.42

Reference	Accession	CM PEPTIDES	CM% COVERAGE by AA	EN PEPTIDE	EN % COVERAGE By AA	ABM PEPTIDES By AA	ABM % COVERAGE By AA
reticulon 4; neuroendocrine-specific protein C like (foocen [Homo sap	24431935					3	8.66
JC7525 chondroitin 4-sulfotransferase (EC 2.8.2.5 - human &	25392157					3	9.24
PKCB_HUMAN Protein kinase C binding protein 1 (Rack7 (Cutaneous T-cell lymph	25453223					4	6.54
similar to /prediction=(method:genefinder, version:084, score:404.95 ~	27481918					2	11.24
similar to hypothetical protein KIAA0404 - human (fragment [Rattus no	27661926					2	9.25
similar to hypothetical protein KIAA0404 - human (fragment [Rattus no	27668709					2	7.25
similar to ubiquitin protein ligase E3A, isoform 1; human papilloma vi	27685325					2	10.21
similar to KIAA1424 protein [imported] - human [	27689359					8	16.81
similar to hypothetical protein DKFZp434B1231.1 - human (fragment [Ra	27712098					2	5.21
similar to ubiquitin specific protease 16; human ubiquitin processing	27714227					4	18.24
S21C_HUMAN Solute carrier family 21 member 12 (Sodium-independent organic ani	27734555					3	11.24
ZAN_HUMAN Zonadhesin precursor	27924006					2	9.81

APPENDIX II List of *falciparum* proteins found in plasma

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
MAL13P1.13 MAL13P1.13 hypothetical protein 17208914:17217659 reverse MW:325177	MAL13P1.13					2	14.31
MAL13P1.133 MAL13P1.133 hypothetical protein 18079812:18096367 reverse MW:658983	MAL13P1.133	7	6.75			8	12.21
MAL13P1.134 MAL13P1.134 helicase, putative 18100324:18104807 reverse MW:130703	MAL13P1.134	2	8.21				
MAL13P1.147 MAL13P1.147 hypothetical protein 18314264:18315493 reverse MW:31757	MAL13P1.147	7	5.43				
MAL13P1.19 MAL13P1.19 hypothetical protein 17261910:17289725 reverse MW:1111825	MAL13P1.19					2	11.21
MAL13P1.201 MAL13P1.201 hypothetical protein 18665462:18667348 forward MW:71608	MAL13P1.201	4	8.24				
MAL13P1.202 MAL13P1.202 hypothetical protein 18667568:18673378 reverse MW:231559	MAL13P1.202	3				2	10.81
MAL13P1.22 MAL13P1.22 DNA ligase 1 17298881:17301785 reverse MW:104506	MAL13P1.22	4	18.75			2	6.31
MAL13P1.221 atcasE aspartate carbamoyltransferase 18869475:18870910 forward MW:43252	MAL13P1.221	2	15.41				
MAL13P1.228 MAL13P1.228 hypothetical protein 18899612:18900850 forward MW:43220	MAL13P1.228	3	8.71				
MAL13P1.23 MAL13P1.23 hypothetical protein 17303750:17305355 reverse MW:57422	MAL13P1.23	4	6.91				
MAL13P1.234 MAL13P1.234 hypothetical protein 18926092:18944453 forward MW:696786	MAL13P1.234	2	5.73			4	7.24
MAL13P1.246 MAL13P1.246 hypothetical protein 19032851:19038762 forward MW:218518	MAL13P1.246					2	8.41

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
MAL13P1.247 MAL13P1.247 hypothetical protein 19040084:19042621 reverse MW:53914	MAL13P1.247					2	6.23
MAL13P1.256 MAL13P1.256 phosphatidylinositol transfer protein, putative 19107032:19112954 forward M	MAL13P1.256	7	12.24	15	20.24	11	17.24
MAL13P1.273 MAL13P1.273 hypothetical protein 19260018:19260314 forward MW:12567	MAL13P1.273	3	18.75				
MAL13P1.273 MAL13P1.273 hypothetical protein 19260018:19260314 forward MW:12567	MAL13P1.278					5	6.51
MAL13P1.32 MAL13P1.32 hypothetical protein 17378053:17390994 forward MW:515254	MAL13P1.32	2	6.54				
MAL13P1.352 MAL13P1.352 hypothetical protein 19843113:19846677 forward MW:131375	MAL13P1.352	2	8.34				
MAL13P1.380 MAL13P1.380 chloroquine resistance marker protein, putative 18807598:18830409 reverse M	MAL13P1.380					5	7.24
MAL13P1.68 MAL13P1.68 hypothetical protein 17645729:17646600 forward MW:25077	MAL13P1.68					2	9.83
MAL7P1.120 MAL7P1.120 hypothetical protein 7621216:7624881 reverse MW:145493	MAL7P1.120	3	9.24			2	12.21
MAL7P1.138 MAL7P1.138 hypothetical protein 7737019:7742007 forward MW:190801	MAL7P1.138					2	6.23
MAL7P1.146 MAL7P1.146 hypothetical protein 7780598:7796259 reverse MW:592976	MAL7P1.146					2	5.37
MAL7P1.15 MAL7P1.15 hypothetical protein 6825470:6838162 reverse MW:513832	MAL7P1.15	2	10.35				
MAL7P1.16 MAL7P1.16 hypothetical protein 6839255:6849061 reverse MW:393472	MAL7P1.16	2	11.25			9	13.47

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
MAL7P1.167 MAL7P1.167 hypothetical protein, conserved 7977659:7985980 reverse MW:329471	MAL7P1.167	16	8.21	8	6.81	24	10.61
MAL7P1.17 MAL7P1.17 hypothetical protein 6851802:6862481 forward MW:425272	MAL7P1.17					2	6.63
MAL7P1.171 MAL7P1.171 hypothetical protein 7997142:8003580 forward MW:244197	MAL7P1.171					2	5.73
MAL7P1.187 VAR erythrocyte membrane protein 1 (PfEMP1) 8089355:8098001 reverse MW:294579	MAL7P1.187					2	6.52
MAL7P1.228 HSP70 Heat Shock 70 KDa Protein, (HSP70) 6722615:6724917 forward MW:73214	MAL7P1.228	3	14.23				
MAL7P1.50 VAR erythrocyte membrane protein 1 (PfEMP1) 7214675:7222289 reverse MW:257012	MAL7P1.50					3	8.17
MAL8P1.113 MAL8P1.113 Peptidase family C50, putative 8599207:8616609 reverse MW:686421	MAL8P1.113	6	16.71	6	16.71	3	7.24
MAL8P1.124 MAL8P1.124 hypothetical protein 8524226:8528314 forward MW:162114	MAL8P1.124					2	6.20
MAL8P1.154 MAL8P1.154 hypothetical protein 8248080:8262068 forward MW:551979	MAL8P1.154					7	9.05
MAL8P1.201 MAL8P1.201 hypothetical protein, conserved 8424105:8427773 reverse MW:128608	MAL8P1.201	2	19.21			2	5.71
MAL8P1.203 MAL8P1.203 hypothetical protein 8433274:8437960 reverse MW:173406	MAL8P1.203					2	8.21
MAL8P1.205 MAL8P1.205 hypothetical protein 9458639:9461485 forward MW:101272	MAL8P1.205	3	18.74			4	16.23
MAL8P1.23 MAL8P1.23 ubiquitin-protein ligase 1, putative 9232549:9258324 reverse MW:1005165	MAL8P1.23					4	12.27
MAL8P1.29 MAL8P1.29 hypothetical protein 9205228:9210428 forward MW:197524	MAL8P1.29	5	12.91				

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
MAL8P1.73 MAL8P1.73 hypothetical protein 8925031:8932699 reverse MW:140512	MAL8P1.73					2	6.24
PF07_0014 PF07_0014 hypothetical protein 6821571:6824747 forward MW:125009	PF07_0014					2	8.21
PF07_0021 PF07_0021 hypothetical protein 6924053:6927022 forward MW:118345	PF07_0021					2	7.15
PF07_0037 PF07_0037 Cg2 protein 7085329:7093518 reverse MW:325496	PF07_0037					4	5.43
PF07_0051 VAR erythrocyte membrane protein 1 (PfEMP1) 7238495:7246077 reverse MW:251678	PF07_0051					3	8.17
PF07_0067 PF07_0067 hypothetical protein 7379466:7382675 reverse MW:126477	PF07_0067					3	7.02
PF07_0107 PF07_0107 hypothetical protein 7765333:7767889 reverse MW:95180	PF07_0107	2	20.51			6	30.21
PF07_0118 PF07_0118 hypothetical protein 7865383:7882068 reverse MW:671128	PF07_0118					3	6.71
PF08_0012 PF08_0012 hypothetical protein 9318038:9325237 forward MW:283552	PF08_0012	3	30.17				
PF08_0028 PF08_0028 hypothetical protein 9194980:9198810 forward MW:154870	PF08_0028					5	8.72
PF08_0080 PF08_0080 hypothetical protein 8784777:8786324 reverse MW:61577	PF08_0080	2	18.61				
PF08_0111 PF08_0111 hypothetical protein 8492131:8496096 reverse MW:157199	PF08_0111	3	7.21			9	13.24
PF08_0140 VAR erythrocyte membrane protein 1 (PfEMP1) 8160064:8170055 forward MW:340615	PF08_0140					3	6.23
PF08_0141 VAR erythrocyte membrane protein 1 (PfEMP1) 8148816:8158149 reverse MW:330699	PF08_0141	2	6.53				

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF10_0001 PF10_0001 erythrocyte membrane protein 1 (PfEMP1) 11107885:11115559 forward MW:252333	PF10_0001	8	7.21			2	5.21
PF10_0021 PF10_0021 hypothetical protein 11171968:11172921 reverse MW:32183	PF10_0021					3	6.21
PF10_0026 PF10_0026 hypothetical protein 11192324:11195410 forward MW:118672	PF10_0026					2	5.73
PF10_0044 PF10_0044 hypothetical protein 11261639:11266495 forward MW:141923	PF10_0044	2	6.25				
PF10_0075 PF10_0075 asparagine-rich antigen 11380520:11385313 reverse MW:182669	PF10_0075					2	8.17
PF10_0117 PF10_0117 hypothetical protein 11542886:11546399 forward MW:94969	PF10_0117					2	7.23
PF10_0140 PF10_0140 hypothetical protein 11645198:11652453 forward MW:283133	PF10_0140	28	8.24			4	6.27
PF10_0165 PF10_0165 DNA polymerase delta catalytic subunit 11764971:11768255 forward MW:126885	PF10_0165	15	12.91				
PF10_0179 PF10_0179 hypothetical protein 11815762:11822799 reverse MW:216365	PF10_0179					4	5.21
PF10_0180 PF10_0180 hypothetical protein 11826421:11828280 forward MW:72959	PF10_0180	3	16.83				
PF10_0189 PF10_0189 hypothetical protein 11875402:11879316 forward MW:154011	PF10_0189					2	5.07
PF10_0215 PF10_0215 hypothetical protein 12004673:12007945 reverse MW:86296	PF10_0215					5	6.17
PF10_0224 PF10_0224 dynein heavy chain, putative 12042889:12059952 reverse MW:671885	PF10_0224	2	17.83			5	28.43
PF10_0241 PF10_0241 hypothetical protein 12109436:12110926 forward MW:59650	PF10_0241					2	6.08

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF10_0251 PF10_0251 hypothetical protein 12152933:12166384 reverse MW:488784	PF10_0251					4	10.84
PF10_0262 PF10_0262 hypothetical protein 12192887:12197965 reverse MW:196494	PF10_0262					2	9.17
PF10_0267 PF10_0267 hypothetical protein 12217392:12218954 forward MW:61468	PF10_0267	11	20.25	9	15.43	19	30.25
PF10_0274 PF10_0274 hypothetical protein 12239747:12242813 forward MW:53268	PF10_0274					3	
PF10_0342 PF10_0342 hypothetical protein 12470836:12472521 forward MW:64879	PF10_0342	2	11.25				
PF10_0403 PF10_0403 rifin 12711791:12713058 forward MW:41245	PF10_0403					2	12.01
PF11_0057 PF11_0057 hypothetical protein 12969380:12971311 forward MW:75901	PF11_0057	3	10.83			3	10.83
PF11_0077 PF11_0077 hypothetical protein 13042775:13046716 forward MW:155821	PF11_0077					2	11.17
PF11_0080 PF11_0080 hypothetical protein 13056716:13057335 forward MW:20354	PF11_0080					2	12.24
PF11_0092 PF11_0092 hypothetical protein 13110962:13116400 reverse MW:213624	PF11_0092	2	6.21				
PF11_0127 PF11_0127 hypothetical protein 13229339:13236721 forward MW:285610	PF11_0127	4	8.73			6	15.68
PF11_0129 PF11_0129 hypothetical protein 13238995:13241723 reverse MW:102084	PF11_0129	3	9.21	5	14.27	5	14.27
PF11_0153 PF11_0153 hypothetical protein 13314431:13316344 reverse MW:77378	PF11_0153					5	18.01
PF11_0158 PF11_0158 hypothetical protein 13329065:13336507 forward MW:295383	PF11_0158					2	14.83

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF11_0185 PF11_0185 hypothetical protein 13445800:13450329 forward MW:180156	PF11_0185	3	12.21				
PF11_0207 PF11_0207 hypothetical protein 13523240:13526567 forward MW:116349	PF11_0207	3	22.41				
PF11_0226 PF11_0226 hypothetical protein 13590287:13596361 reverse MW:238901	PF11_0226	2	23.22			3	24.59
PF11_0240 PF11_0240 dynein heavy chain, putative 13646017:13662321 forward MW:617391	PF11_0240	16	16.51			2	5.57
PF11_0264 PF11_0264 DNA-dependent RNA polymerase 13758911:13763506 forward MW:179993	PF11_0264					2	6.71
PF11_0270 PF11_0270 threonine -- tRNA ligase, putative 13781382:13784423 reverse MW:119546	PF11_0270	2	12.31				
PF11_0297 PF11_0297 hypothetical protein 13877483:13879162 reverse MW:65782	PF11_0297	3	14.21				
PF11_0326 PF11_0326 hypothetical protein 13983568:13991859 forward MW:330308	PF11_0326	5	18.31			6	19.28
PF11_0347 PF11_0347 hypothetical protein 14069220:14073034 forward MW:120540	PF11_0347					6	20.75
PF11_0374 PF11_0374 hypothetical protein 14189790:14193247 forward MW:128546	PF11_0374	4	10.21	2	5.17	4	10.21
PF11_0392 PF11_0392 hypothetical protein 14254640:14263189 forward MW:335925	PF11_0392					5	11.27
PF11_0404 PF11_0404 malaria antigen 14323790:14332091 reverse MW:309447	PF11_0404	2	16.51			2	16.51
PF11_0408 PF11_0408 hypothetical protein 14352439:14359080 reverse MW:265269	PF11_0408					3	12.13
PF11_0417 PF11_0417 hypothetical protein 14383462:14388936 reverse MW:219858	PF11_0417	4	14.23				

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF11_0468 PF11_0468 hypothetical protein 14580463:14581617 forward MW:46253	PF11_0468			2	5.57	3	8.31
PF11_0481 PF11_0481 hypothetical protein 14640193:14648460 forward MW:325331	PF11_0481					5	9.61
PF11_0511 PF11_0511 hypothetical protein 14752916:14753482 reverse MW:22912	PF11_0511	4	8.71			4	28.71
PF11_0511 PF11_0511 hypothetical protein 14752916:14753482 reverse MW:22912	PF11_0511						
PF11_0528 PF11_0528 hypothetical protein 13259139:13277786 forward MW:710233	PF11_0528	6	6.63			5	5.17
PF11_0540 PF11_0540 hypothetical protein 13149618:13156501 reverse MW:218787	PF11_0540					5	8.21
PF13_0035 PF13_0035 hypothetical protein 17404959:17407994 forward MW:120726	PF13_0035					5	8.07
PF13_0060 PF13_0060 hypothetical protein 17556418:17557905 forward MW:61233	PF13_0060	2	6.21				
PF13_0072 PF13_0072 hypothetical protein 17613767:17621588 reverse MW:277699	PF13_0072					4	11.71
PF13_0078 PF13_0078 hypothetical protein 17650234:17655544 reverse MW:209078	PF13_0078					2	5.81
PF13_0126 PF13_0126 hypothetical protein 18021339:18026349 reverse MW:178351	PF13_0126					2	7.71
PF13_0161 PF13_0161 hypothetical protein 18222783:18228170 forward MW:205479	PF13_0161					2	6.51
PF13_0167 PF13_0167 hypothetical protein 18383715:18385060 forward MW:48617	PF13_0167					2	6.41
PF13_0172 PF13_0172 hypothetical protein 18406552:18408120 forward MW:61248	PF13_0172					2	5.73

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF13_0189 PF13_0189 hypothetical protein 18469675:18472626 reverse MW:117790	PF13_0189	3	14.12				
PF13_0239 PF13_0239 hypothetical protein 18864233:18867967 reverse MW:150359	PF13_0239					2	5.28
PF13_0354 PF13_0354 alanine--tRNA ligase, putative 19768852:19773078 forward MW:165137	PF13_0354	4	13.17			2	8.71
PF13_0355 PF13_0355 hypothetical protein 19774425:19777973 forward MW:135835	PF13_0355	3	16.23				
PF14_0059 PF14_0059 hypothetical protein 20190313:20198775 forward MW:338385	PF14_0059					3	7.98
PF14_0073 PF14_0073 hypothetical protein 20243482:20249583 reverse MW:241134	PF14_0073					2	6.88
PF14_0074 PF14_0074 hypothetical protein 20251644:20253141 forward MW:47264	PF14_0074	2	18.21				
PF14_0081 PF14_0081 DNA repair helicase, putative 20286443:20289925 reverse MW:135547	PF14_0081	6	6.23	12	10.41	10	15.62
PF14_0082 PF14_0082 hypothetical protein 20292742:20293329 forward MW:22965	PF14_0082					3	8.88
PF14_0084 PF14_0084 hypothetical protein 20297222:20318936 reverse MW:847899	PF14_0084					5	6.51
PF14_0088 PF14_0088 hypothetical protein 20329186:20332656 forward MW:136528	PF14_0088	2	14.23				
PF14_0101 PF14_0101 hypothetical protein 20376120:20389530 reverse MW:526740	PF14_0101	4	10.23				
PF14_0102 PF14_0102 rhoptry-associated protein 1 20392910:20395258 forward MW:90053	PF14_0102	2	11.24				
PF14_0112 PF14_0112 POM1, putative 20433869:20439919 reverse MW:235822	PF14_0112	2	10.24			2	7.21

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF14_0155 PF14_0155 hypothetical protein 20607282:20609000 reverse MW:65533	PF14_0155					4	15.62
PF14_0165 PF14_0165 hypothetical protein 20642159:20651239 reverse MW:357642	PF14_0165					2	6.17
PF14_0170 PF14_0170 hypothetical protein 20667052:20680644 forward MW:519836	PF14_0170					2	5.08
PF14_0175 PF14_0175 hypothetical protein 20703161:20717149 reverse MW:548876	PF14_0175					3	6.23
PF14_0190 PF14_0190 hypothetical protein 20787788:20790709 reverse MW:54743	PF14_0190					2	6.43
PF14_0202 PF14_0202 dynein-associated protein, putative 20836037:20837234 reverse MW:30842	PF14_0202	6	7.24				
PF14_0263 PF14_0263 hypothetical protein 21081555:21087880 reverse MW:224359	PF14_0263					2	7.21
PF14_0277 PF14_0277 coatamer protein, beta subunit, putative 21143239:21147703 forward MW:162835	PF14_0277					2	5.25
PF14_0281 PF14_0281 aspartyl protease, putative 21160759:21163876 forward MW:74183	PF14_0281	2	6.23				
PF14_0282 PF14_0282 hypothetical protein, conserved 21164356:21172329 reverse MW:306811	PF14_0282					3	6.41
PF14_0293 PF14_0293 hypothetical protein 21206385:21209533 reverse MW:116331	PF14_0293			10	12.41		
PF14_0318 PF14_0318 hypothetical protein 21315424:21321405 reverse MW:243379	PF14_0318					2	10.41
PF14_0326 PF14_0326 hypothetical protein 21350759:21375076 reverse MW:969736	PF14_0326	9	12.47			3	5.17

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF14_0334 PF14_0334 NAD(P)H-dependent glutamate synthase, putative 21394312:21403632 reverse MW:353	PF14_0334	2	11.22				
PF14_0345 PF14_0345 hypothetical protein 21456577:21462412 forward MW:201224	PF14_0345	3	19.87				
PF14_0364 PF14_0364 cleavage and polyadenylation specifity factor protein, putative 21529938:215325	PF14_0364	139	30.34				
PF14_0372 PF14_0372 hypothetical protein 21566916:21572918 forward MW:221690	PF14_0372	8	5.74	23	13.28	13	10.88
PF14_0372 PF14_0372 hypothetical protein 21566916:21572918 forward MW:221690	PF14_0372						
PF14_0379 PF14_0379 hypothetical protein 21596598:21603778 reverse MW:259038	PF14_0379					5	10.21
PF14_0383 PF14_0383 hypothetical protein 21620084:21624105 reverse MW:154203	PF14_0383	4	7.89	4	7.89	6	6.24
PF14_0392 PF14_0392 Ser/Thr protein kinase, putative 21654202:21661568 forward MW:263797	PF14_0392					10	8.88
PF14_0402 PF14_0402 hypothetical protein 21689582:21695923 forward MW:249607	PF14_0402	5	10.39				
PF14_0404 PF14_0404 hypothetical protein 21702813:21713327 forward MW:408313	PF14_0404					6	5.67
PF14_0405 PF14_0405 hypothetical protein 21714524:21722201 forward MW:300985	PF14_0405					4	7.21
PF14_0417 PF14_0417 heat shock protein, putative 21769699:21772482 forward MW:107000	PF14_0417					2	5.03
PF14_0419 PF14_0419 hypothetical protein 21774372:21797737 reverse MW:857945	PF14_0419	10	11.29			4	5.51
PF14_0435 PF14_0435 hypothetical protein 21851069:21854988 reverse MW:144113	PF14_0435	9	14.34			20	20.25

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF14_0454 PF14_0454 hypothetical protein 21919657:21925617 forward MW:238053	PF14_0454	3	12.24			4	14.47
PF14_0470 PF14_0470 hypothetical protein 21992620:21997116 reverse MW:152079	PF14_0470					2	5.43
PF14_0473 PF14_0473 3'-5' exonuclease, putative 22010537:22013947 reverse MW:135629	PF14_0473	2	4.56			3	6.21
PF14_0478 PF14_0478 hypothetical protein 22029046:22031643 forward MW:103074	PF14_0478					2	5.41
PF14_0480 PF14_0480 hypothetical protein 22040144:22046061 reverse MW:199741	PF14_0480	2	8.74				
PF14_0490 PF14_0490 hypothetical protein 22085965:22086408 reverse MW:17257	PF14_0490					2	5.32
PF14_0504 PF14_0504 hypothetical protein 22144273:22149116 reverse MW:144251	PF14_0504	3	9.21				
PF14_0507 PF14_0507 hypothetical protein 22160582:22166353 forward MW:233304	PF14_0507					2	6.71
PF14_0509 PF14_0509 hypothetical protein 22171773:22177110 reverse MW:208230	PF14_0509			4	6.24		
PF14_0530 PF14_0530 ferlin, putative 22243933:22249647 reverse MW:224724	PF14_0530					2	7.20
PF14_0538 PF14_0538 hypothetical protein 22286784:22293959 forward MW:283206	PF14_0538	9	6.47			10	7.17
PF14_0626 PF14_0626 dynein beta chain, putative 22634424:22655109 forward MW:771768	PF14_0626	4	5.21			9	8.17
PF14_0631 PF14_0631 hypothetical protein 22667217:22676453 reverse MW:355943	PF14_0631					2	10.24
PF14_0668 PF14_0668 hypothetical protein 22842473:22849618 forward MW:286361	PF14_0668					2	6.72

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFA0030c RIF RIFIN 54001:55229 reverse MW:40531	PFA0030c					2	6.43
PFA0100c PFA0100c hypothetical protein 93346:94066 reverse MW:22895	PFA0100c	6	11.24	13	20.88	11	15.61
PFA0180w PFA0180w hypothetical protein 161365:166464 forward MW:175729	PFA0180w	4	10.91			6	15.71
PFA0185w PFA0185w hypothetical protein 166748:168632 forward MW:62168	PFA0185w	3	8.73			2	5.17
PFA0280w PFA0280w asparagine-rich antigen Pfa35-2 238905:248096 forward MW:364077	PFA0280w					5	10.24
PFA0290w PFA0290w DNA binding protein, putative 253399:255169 forward MW:54941	PFA0290w			3	6.75		
PFA0320w PFA0320w hypothetical protein 274530:278765 forward MW:167054	PFA0320w	3	6.75				
PFA0345w PFA0345w centrin, putative 293860:295186 forward MW:19599	PFA0345w	2	6.42				
PFA0550w PFA0550w hypothetical protein 437026:439716 forward MW:99011	PFA0550w					6	8.22
PFA0625w PFA0625w hypothetical protein 496393:501225 forward MW:192352	PFA0625w					3	9.51
PFA0665w PFA0665w hypothetical protein 528829:538073 forward MW:346568	PFA0665w					2	9.23
PFA0735w PFA0735w hypothetical protein 589111:590114 forward MW:34835	PFA0735w					2	6.54
PFB0010w PFB0010w erythrocyte membrane protein 1 (PfEMP1) 668524:674460 forward MW:196552	PFB0010w	2	8.21				
PFB0095c PFB0095c erythrocyte membrane protein 3 734610:742130 reverse MW:273651	PFB0095c	4	8.92			5	12.10

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFB0130w PFB0130w polyprenyl synthetase, putative 778815:780431 forward MW:63046	PFB0130w					2	5.07
PFB0150c PFB0150c protein kinase, putative 792816:802952 reverse MW:352872	PFB0150c	3	5.67			5	6.17
PFB0205c PFB0205c hypothetical protein 843774:847319 reverse MW:142006	PFB0205c	12	12.23				
PFB0350c PFB0350c cysteine protease, putative 956741:960033 reverse MW:105423	PFB0350c					3	7.07
PFB0405w PFB0405w transmission-blocking target antigen s230 precursor 1013730:1023137 forward MW:36	PFB0405w	5	5.71				
PFB0555c PFB0555c hypothetical protein 1138458:1151146 reverse MW:494221	PFB0555c					4	6.83
PFB0665w PFB0665w Ser/Thr protein kinase, putative 1237488:1242632 forward MW:204291	PFB0665w	2	7.21				
PFB0760w PFB0760w hypothetical protein 1329776:1331836 forward MW:77594	PFB0760w					2	7.27
PFB0770c PFB0770c hypothetical protein 1339484:1342606 reverse MW:122455	PFB0770c	8	8.27			12	15.32
PFB0775w PFB0775w hypothetical protein 1344536:1346185 forward MW:64574	PFB0775w	2	6.23			2	7.21
PFB0925w PFB0925w hypothetical protein 1463719:1465831 forward MW:71820	PFB0925w	2	5.41				
PFB1055c PFB1055c erythrocyte membrane protein 1 (PfEMP1) 1559644:1566940 reverse MW:249957	PFB1055c	2	5.39	4	9.07	3	7.17
PFC0060c PFC0060c Serine/threonine protein kinase, putative 1665267:1667465 reverse MW:70533	PFC0060c			3	10.14		

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFC0120w PFC0120w Cytoadherence linked asexual protein, CLAG 1722491:1727733 forward MW:167243	PFC0120w					11	12.25
PFC0165w PFC0165w hypothetical protein, conserved 1769801:1776481 forward MW:267980	PFC0165w					3	5.78
PFC0221c PFC0221c hypothetical protein, conserved 1818861:1819373 reverse MW:20838	PFC0221c					2	5.15
PFC0335c PFC0335c hypothetical protein 1927097:1938721 reverse MW:448215	PFC0335c					2	6.31
PFC0425w PFC0425w hypothetical protein 2014358:2028010 forward MW:534122	PFC0425w					2	6.51
PFC0485w PFC0485w putative protein kinase 2075874:2083580 forward MW:295501	PFC0485w					2	7.24
PFC0615w PFC0615w hypothetical protein 2191667:2195000 forward MW:108634	PFC0615w	3	6.83			2	6.87
PFC0705c PFC0705c hypothetical protein 2236917:2244315 reverse MW:293717	PFC0705c					2	5.25
PFC0755c PFC0755c protein kinase, putative 2277221:2281882 reverse MW:182220	PFC0755c	8	7.21			5	6.32
PFC0760c PFC0760c hypothetical protein 2286609:2296793 reverse MW:402958	PFC0760c	2	6.51				
PFC0805w PFC0805w DNA-directed RNA polymerase II, putative 2335530:2342903 forward MW:278679	PFC0805w	2	5.43				
PFC0810c PFC0810c hypothetical protein 2343556:2347357 reverse MW:132984	PFC0810c	3	5.73				
PFC0870w PFC0870w elongation factor 1 (EF-1), putative 2404977:2405759 forward MW:17706	PFC0870w	3	5.87				
PFC0890w PFC0890w vesicle transport protein, putative 2427426:2428822 forward MW:25934	PFC0890w	2	2.41			4	5.61

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFC0905c PFC0905c hypothetical protein 2435311:2444568 reverse MW:360528	PFC0905c	2	6.51			3	7.45
PFC0960c PFC0960c hypothetical protein 2490827:2497181 reverse MW:231796	PFC0960c	3	5.71			9	16.21
PFD0050w RIF RIFIN 2736121:2737379 forward MW:41001	PFD0050w					3	9.07
PFD0135c RIF truncated RIFIN, truncated 2819955:2820669 reverse MW:18243	PFD0135c	2	9.21			2	9.21
PFD0160w PFD0160w hypothetical protein 2839810:2851652 forward MW:455479	PFD0160w	4	8.34			6	9.88
PFD0190w PFD0190w hypothetical protein 2876031:2879782 forward MW:142739	PFD0190w	5	6.71				
PFD0200c PFD0200c hypothetical protein 2881723:2888678 reverse MW:271093	PFD0200c					5	12.11
PFD0320c PFD0320c hypothetical protein 2991240:3001453 reverse MW:401571	PFD0320c	4	5.24			6	8.28
PFD0430c PFD0430c hypothetical protein 3074750:3078202 reverse MW:94516	PFD0430c					2	7.17
PFD0440w PFD0440w hypothetical protein 3081678:3084572 forward MW:80851	PFD0440w	2	7.24				
PFD0462w PFD0462w DNAJ protein 3103378:3105396 forward MW:75916	PFD0462w					2	6.31
PFD0535w PFD0535w hypothetical protein 3143210:3147106 forward MW:150271	PFD0535w	7	9.41			14	15.69
PFD0685c PFD0685c chromosome associated protein, putative 3284355:3290588 reverse MW:141227	PFD0685c	5	16.71				
PFD0740w PFD0740w Plasmodium falciparum protein kinase, putative 3339274:3343432 forward MW:159560	PFD0740w	2	12.43			3	15.71

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFD0790c PFD0790c DNA replication licensing factor, putative 3368556:3372953 reverse MW:171315	PFD0790c	3	8.41				
PFD0805w PFD0805w prohibitin-like protein, putative 3383729:3384592 forward MW:34624	PFD0805w					2	6.12
PFD0840w MAL4P1.163 hypothetical protein 3413756:3434065 forward MW:822401	PFD0840w					2	6.07
PFD0900w PFD0900w hypothetical protein 3491855:3497890 forward MW:236360	PFD0900w	2	7.81				
PFD0935c PFD0935c hypothetical protein 3513359:3514510 reverse MW:41785	PFD0935c					2	5.47
PFD0980w PFD0980w holo-(acyl-carrier protein) synthase, putative 3567649:3569433 forward MW:70386	PFD0980w	3	10.82	2	7.21	3	10.82
PFD0995c VAR erythrocyte membrane protein 1 (PfEMP1) 3589951:3596795 reverse MW:247612	PFD0995c	2	9.43				
PFD1005c VAR erythrocyte membrane protein 1 (PfEMP1) 3612987:3620531 reverse MW:248761	PFD1005c	2	7.64			3	11.17
PFD1045c PFD1045c erythrocyte membrane-associated antigen, putative 3659383:3672168 reverse MW:5085	PFD1045c	2	5.43				
PFD1050w PFD1050w alpha-tubulin ii 3677408:3679087 forward MW:49691	PFD1050w	6	8.41				
PFD1160w PFD1160w hypothetical protein 3754732:3762032 forward MW:286418	PFD1160w	4	7.63				
PFD1235w VAR erythrocyte membrane protein 1 (PfEMP1) 3810526:3821924 forward MW:405213	PFD1235w					2	7.24
PFE0085c PFE0085c hypothetical protein 3941413:3943794 reverse MW:95432	PFE0085c	5	6.71				

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFE0205w PFE0205w ATP-dependent helicase, putative 4022242:4024095 forward MW:59399	PFE0205w	3	5.54			6	14.25
PFE0230w PFE0230w hypothetical protein 4035014:4042063 forward MW:291041	PFE0230w					2	7.18
PFE0240c PFE0240c hypothetical protein 4042406:4059325 reverse MW:672034	PFE0240c	2	6.72				
PFE0360c PFE0360c hypothetical protein 4156523:4159204 reverse MW:107040	PFE0360c					3	5.72
PFE0450w PFE0450w chromosome condensation protein, putative 4235670:4240796 forward MW:201670	PFE0450w					2	5.11
PFE0465c PFE0465c RNA polymerase I 4244507:4253590 reverse MW:340681	PFE0465c					3	10.98
PFE0640w PFE0640w hypothetical protein 4408581:4409378 forward MW:31952	PFE0640w					2	5.24
PFE0890c PFE0890c hypothetical protein 4600389:4601225 reverse MW:26807	PFE0890c					2	6.83
PFE1070c PFE1070c hypothetical protein 4724774:4729160 reverse MW:161905	PFE1070c	11	10.81			2	5.07
PFE1120w PFE1120w hypothetical protein 4762425:4790903 forward MW:1121416	PFE1120w					2	11.03
PFE1255w PFE1255w hypothetical protein 4900051:4905842 forward MW:217231	PFE1255w					3	10.21
PFE1350c PFE1350c ubiquitin-conjugating enzyme, putative 4984547:4985342 reverse MW:17411	PFE1350c					2	5.64
PFE1545c PFE1545c diaphanous homolog, putative 5112309:5120336 reverse MW:309480	PFE1545c					2	6.28

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFE1640w VAR-like erythrocyte membrane protein 1 (PfEMP1), truncated 5188058:5197552 forward MW:367	PFE1640w	2	11.24				
PFF0130c PFF0130c hypothetical protein 5311966:5313366 reverse MW:56493	PFF0130c					10	15.87
PFF0145w PFF0145w hypothetical protein 5320776:5322554 forward MW:69666	PFF0145w					2	5.46
PFF0195c PFF0195c hypothetical protein 5361970:5368002 reverse MW:239916	PFF0195c					2	7.82
PFF0470w PFF0470w hypothetical protein 5600887:5605383 forward MW:124609	PFF0470w					2	5.44
PFF0485c PFF0485c hypothetical protein, conserved 5619297:5620762 reverse MW:33146	PFF0485c	3	10.25				
PFF0670w PFF0670w hypothetical protein 5764279:5777133 forward MW:483835	PFF0670w	2	6.72			2	6.72
PFF0720w PFF0720w hypothetical protein 5815902:5819676 forward MW:130935	PFF0720w	3	8.41				
PFF0750w PFF0750w cdc2-like protein kinase, putative 5842193:5844310 forward MW:82948	PFF0750w					3	10.11
PFF0990c PFF0990c hypothetical protein 6046547:6049006 reverse MW:98972	PFF0990c	5	12.41			6	14.44
PFF1055c PFF1055c hypothetical protein, conserved 6091593:6093923 reverse MW:92453	PFF1055c					3	16.55
PFF1120c PFF1120c hypothetical protein 6145742:6146790 reverse MW:23993	PFF1120c		5.62			2	9.17
PFF1135w PFF1135w transcription or splicing factor-like protein, putative 6154911:6156902 forward M	PFF1135w	3				2	4.02

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFF1140c PFF1140c ATP-dependent DEAD box helicase, putative 6157278:6160691 reverse MW:136287	PFF1140c	3	6.32				
PFF1185w PFF1185w iswi protein homologue 6197670:6205829 forward MW:315624	PFF1185w					2	5.73
PFF1265w PFF1265w oxidoreductase, short-chain dehydrogenase family, putative 6254198:6256675 forwar	PFF1265w	3	7.24			2	8.25
PFF1280w PFF1280w hypothetical protein 6263293:6266772 forward MW:136291	PFF1280w					3	9.64
PFF1355w PFF1355w hypothetical protein 6317733:6319004 forward MW:36635	PFF1355w					2	11.24
PFF1365c PFF1365c hypothetical protein 6320953:6351810 reverse MW:1205926	PFF1365c	5	8.31			4	6.28
PFF1430c PFF1430c transmembrane amino acid transporter protein, putative 6412095:6414152 reverse MW	PFF1430c	2	6.84				
PFF1440w PFF1440w SET-domain protein, putative 6420088:6441070 forward MW:797045	PFF1440w	5	7.21			4	5.67
PFF1450w PFF1450w sec14-like cytosolic factor or phosphatidylinositol/phosphatidylcholine transfer	PFF1450w	5	10.31				
PFF1470c PFF1470c DNA polymerase epsilon, catalytic subunit a, putative 6458546:6467744 reverse MW:	PFF1470c					5	5.54
PFF1490w PFF1490w hypothetical protein 6483206:6485179 forward MW:78441	PFF1490w			4	5.71	4	5.71
PFI0005w VAR erythrocyte membrane protein 1 (PfEMP1) 9557751:9565556 forward MW:258412	PFI0005w					2	5.03

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFI0055c RIF RIFIN 9599013:9600224 reverse MW:38834	PFI0055c					2	5.45
PFI0260c PFI0260c hypothetical protein 9789022:9807378 reverse MW:720597	PFI0260c	5	11.24			4	10.17
PFI0295c PFI0295c hypothetical protein, conserved 9836333:9841847 reverse MW:215079	PFI0295c	2	13.41			2	13.41
PFI0495w PFI0495w hypothetical protein 10002221:10011430 forward MW:368349	PFI0495w					2	7.65
PFI0665w PFI0665w hypothetical protein 10122388:10126083 forward MW:147869	PFI0665w	2	14.21	7	22.41	2	14.21
PFI0700c PFI0700c Met-10	PFI0700c					2	5.73
PFI0805w PFI0805w hypothetical protein 10221224:10228783 forward MW:303756	PFI0805w					2	5.86
PFI0960w PFI0960w dolichyl-diphosphooligosaccharide--protein-glyc ot ransferase,putative 10338364:1	PFI0960w					3	6.83
PFI0975c PFI0975c hypothetical protein 10347320:10357465 reverse MW:396158	PFI0975c			3	7.21	2	6.33
PFI1150w PFI1150w hypothetical protein 10488069:10491383 forward MW:135820	PFI1150w					2	5.21
PFI1265w PFI1265w hypothetical protein 10570976:10576240 forward MW:208593	PFI1265w	3	16.71			2	14.28
PFI1305w PFI1305w hypothetical protein 10622357:10626456 forward MW:156852	PFI1305w	2	12.73				
PFI1385c PFI1385c hypothetical protein 10677601:10680048 reverse MW:83024	PFI1385c	5	11.24			2	5.17
PFI1410c PFI1410c hypothetical protein 10688309:10692814 reverse MW:181365	PFI1410c					2	5.02

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFI1500w PFI1500w hypothetical protein 10762652:10777007 forward MW:542006	PFI1500w	2	14.71	3	16.21	3	16.21
PFL0015c RIF RIFIN 14835464:14836647 reverse MW:38099	PFL0015c	6	6.32			9	18.47
PFL0045c 2277.t00009 hypothetical protein 14878615:14879908 reverse MW:46247	PFL0045c					3	7.81
PFL0130c 2277.t00026 hypothetical protein, conserved 14953623:14960862 reverse MW:251678	PFL0130c					2	5.63
PFL0405w 2277.t00082 hypothetical protein 15167281:15181255 forward MW:530159	PFL0405w	2	12.21				
PFL0410w 2277.t00083 hypothetical protein 15181933:15192864 forward MW:418854	PFL0410w	6	5.67				
PFL0440c 2277.t00089 hypothetical protein 15210989:15214531 reverse MW:138956	PFL0440c	4	8.32			4	8.32
PFL0555c 2277.t00112 hypothetical protein 15297272:15302767 reverse MW:219651	PFL0555c	5	12.21			5	12.21
PFL0610w 2277.t00123 hypothetical protein 15348898:15353912 forward MW:146965	PFL0610w					2	5.07
PFL0650c PFL0650c hypothetical protein conserved 15384974:15386704 reverse MW:66006	PFL0650c	2	14.21				
PFL0770w 2277.t00155 seryl-tRNA synthetase, putative 15449918:15451774 forward MW:73273	PFL0770w	2	15.21				
PFL0785c 2277.t00158 signal recognition particle 19 kd protein, putative 15456777:15457235 reverse	PFL0785c	7	13.20				
PFL1035w 2277.t00207 hypothetical protein 15665641:15668169 forward MW:99321	PFL1035w					2	6.83
PFL1110c 2277.t00222 cAMP-dependent protein kinase regulatory subunit, putative 15733202:15735368 r	PFL1110c					2	5.07

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFL1170w 2277.t00234 polyadenylate-binding protein, putative 15794005:15796632 forward MW:97230	PFL1170w					2	5.17
PFL1280w 2277.t00256 hypothetical protein 15877835:15882352 forward MW:179854	PFL1280w					2	5.45
PFL1320w 2277.t00264 hypothetical protein 15910627:15914115 forward MW:141025	PFL1320w	3	22.21				
PFL1330c 2277.t00266 hypothetical protein 15919401:15926469 reverse MW:273651	PFL1330c	13	20.80	5	5.61	15	21.07
PFL1395c 2277.t00279 hypothetical protein 15974933:15984562 reverse MW:378915	PFL1395c					3	10.24
PFL1410c 2277.t00282 hypothetical protein 15998263:16004589 reverse MW:248306	PFL1410c					2	7.83
PFL1525c 2277.t00305 pre-mRNA splicing factor RNA helicase, putative 16106474:16109980 reverse MW:1	PFL1525c					2	5.67
PFL1540c 2277.t00308 phenylalanyl-tRNA synthetase alpha chain, putative 16131126:16132715 reverse M	PFL1540c					2	8.24
PFL1560c 2277.t00312 hypothetical protein 16143085:16144191 reverse MW:44070	PFL1560c					2	5.71
PFL1645w 2277.t00329 hypothetical protein 16219787:16231477 forward MW:466023	PFL1645w	4	15.61			2	7.24
PFL1650w 2277.t00330 hypothetical protein 16232771:16238973 forward MW:242182	PFL1650w					4	9.24
PFL1675c 2277.t00335 hypothetical protein 16251748:16255708 reverse MW:147891	PFL1675c	3	14.32				
PFL1930w 2277.t00386 hypothetical protein 16465825:16483128 forward MW:691427	PFL1930w					5	12.25

List of falciparum proteins found in plasma Cont'd

Reference	Accession	CM Peptides	CM% Coverage by AA	EN Peptides	EN% Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFL1940w 2277.t00388 3-hydroxyisobutyryl-coenzyme A hydrolase, putative 16488209:16489834 forward M	PFL1940w	2	14.47			3	16.78
PFL1950w VAR erythrocyte membrane protein 1 (PfEMP1) 16499524:16508461 forward MW:300289	PFL1950w	4	12.20				
PFL2085w 2277.t00417 hypothetical protein 16624111:16625442 forward MW:53660	PFL2085w	2	11.21				
PFL2110c 2277.t00422 hypothetical protein 16639438:16646110 reverse MW:221639	PFL2110c					2	7.07
PFL2150c 2277.t00430 hypothetical protein, conserved 16678568:16680661 reverse MW:39692	PFL2150c	5	16.20			3	10.27
PFL2295w 2277.t00459 hypothetical protein 16777810:16778448 forward MW:25668	PFL2295w					2	5.07
PFL2350c 2277.t00470 hypothetical protein 16810547:16812565 reverse MW:81544	PFL2350c	27	14.21				
PFL2375c 2277.t00475 cutA, putative 16822736:16823841 reverse MW:18685	PFL2375c					5	6.24
PFL2390c 2277.t00478 hypothetical protein 16835375:16845223 reverse MW:348683	PFL2390c					3	9.21
PFL2520w 2277.t00502 reticulocyte-binding protein, putative 16941423:16950244 forward MW:328954	PFL2520w					2	10.24
PFL2665c VAR erythrocyte membrane protein 1 (PfEMP1) 17046640:17054331 reverse MW:255417	PFL2665c	2	11.21			2	11.21

APPENDIX III List of human proteins found in CSF

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
C28894 myeloperoxidase (EC 1.11.1.7), splice form H14 - human &	88180	-	-	-	-	9	12.73
B34611 3',5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35) alpha chain - human	105117	2	6.75	-	-	4	5.94
AACT_HUMAN Alpha-1-antichymotrypsin precursor (ACT)	112874	-	-	-	-	4	8.75
ALC1_HUMAN Ig alpha-1 chain C region	113584	-	-	-	-	3	7.93
ARY1_HUMAN Arylamine N-acetyltransferase 1 (Arylamide acetylase 1) (Arylamine N	114234	-	-	1	4.83	2	9.31
K2C3_HUMAN Keratin, type II cytoskeletal 3 (Cytokeratin 3) (K3) (CK3) (65 kDa c	125098	5	9.86	6	7.47	5	8.43
CH60_HUMAN 60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa c	129379	-	-	-	-	-	1.57
PGH1_HUMAN Prostaglandin G/H synthase 1 precursor (Cyclooxygenase -1) (COX-1) (	129899	-	-	-	-	4	7.68
SN21_HUMAN Possible global transcription activator SNF2L1	134584	-	-	-	-	16	17.42
sperm activating protein subunit I, apolipoprotein A1, SPAP subunit I [hum	235865	-	-	-	-	2	50.00
PRHU3 proteinase 3 (EC 3.4.21.-) precursor [validated] – human	279563	-	-	-	-	5	16.80
A45023 lamin B2 - human (fragment)	345758	-	-	-	-	3	6.41
A Chain A, Transthyretin (Also Called Prealbumin) Complex With Milrinone	494652	2	14.96	4	25.98	2	20.4
A53041 effector cell proteinase receptor 1 - human	1082345	-	-	-	-	4	10.39

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
K2C1_HUMAN Keratin, type II cytoskeletal 1 (Cytokeratin 1) (K1) (CK 1) (67 kDa	1346343	6	9.94	14	22.05	4	28.77
MYHA_HUMAN Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,	1346640	-	-	-	-	8	4.55
Similar to Human C219-reactive peptide (L34688) [H	1665825	-	-	-	-	8	8.13
A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1	1703025	-	-	4	10.04	14	26.08
I53799 CG1 protein - human	2134903	-	-	-	-	8	6.31
CAA85776. I38344 titin, cardiac muscle [validated] - human family, mem	2136280	-	-	-	-	110	4.12
S69339 Ig heavy chain V region precursor - human iciency viru	2146957	-	-	4	11.76	6	14.44
A Chain A, Human Cathepsin G	2392230	-	-	-	-	3	13.39
A Chain A, Crystal Structure Of Human Apolipoprotein A-I	2914175	-	-	-	-	9	49.25
B Chain B, Crystal Structure Of S-Nitroso-Nitrosyl Human Hemoglobin A	3660145	1	8.90	4	28.77	4	28.77
C Chain C, Human Serum Transferrin, Recombinant N-Terminal Lobe, Apo Form	4389232	6	20.18	4	13.95	5	16.62
Human Serum Albumin In A Complex With Myristic Acid And Tri-Iodobenzoic Acid	4389275	20	28.57	20	27.37	24	35.28
copper containing amine oxidase 3 precursor; amine oxidase (copper-cont	4502119	-	-	-	-	2	2.88
centromere protein E; Centromere autoantigen E (3	4502781	-	-	-	-	31	11.57
gelsolin (amyloidosis, Finnish type); Gelsolin [Homo sapiens]	4504165	-	-	-	-	2	2.56

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
guanylate cyclase 2F; RetGC-2; guanylate cyclase 2D-like, membrane (ret	4504219	6	4.78	-	-	-	-
alpha 2 globin [Homo sapiens]	4504345	-	-	2	12.68	2	12.68
delta globin [Homo sapiens]	4504351	3	14.29	3	14.29	3	21.09
A-gamma globin; hemoglobin, gamma A; hemoglobin gamma-a chain [Homo sap	4504353	-	-	4	20.84	9	50.34
H factor 1 (complement); H factor-1 (complement) [Homo sapiens]	4504375	-	-	-	-	5	3.41
orosomucoid 2; alpha-1-acid glycoprotein, type 2	4505529	-	-	-	-	4	15.92
S100 calcium-binding protein A9; calgranulin B [Homo sapiens]	4506773	-	-	-	-	5	36.84
secretogranin II precursor; Chromogranin C (secre	4506801	3	5.67	-	-	-	-
solute carrier family 5 (sodium/glucose cotransporter), member 1; Human	4507031	2	8.4	-	-	-	-
sulfotransferase family, cytosolic, 1C, member 1; sulfotransferase 1C1;	4507305	-	-	2	8.78	-	-
thyrotrophic embryonic factor; Thyrotroph embryon	4507431	-	-	-	-	5	12.54
vav 2 oncogene; Oncogene VAV2 [Homo sapiens]	4507871	-	-	-	-	6	7.52
wingless-type MMTV integration site family, member 8B precursor [Homo s	4507931	-	-	3	7.12	-	-
zinc finger protein 91 (HPF7, HTF10) [Homo sapien	4508041	-	-	-	-	5	4.87
alpha 2 macroglobulin precursor [Homo sapiens]	4557225	-	-	-	-	15	10.52

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
alanine-glyoxylate aminotransferase; alanine-glyoxylate aminotransferas	4557289	3	8.93	-	-	-	-
apolipoprotein A-I precursor [Homo sapiens]	4557321	-	-	-	-	4	4.87
butyrylcholinesterase precursor [Homo sapiens]	4557351	-	-	-	-	3	6.31
Bruton agammaglobulinemia tyrosine kinase [Homo s	4557377	-	-	-	-	4	7.28
complement component 3 precursor [Homo sapiens]	4557385	-	-	-	-	17	10.70
jagged 1 precursor; jagged 1; jagged1 (Alagille syndrome) [Homo sapiens	4557679	-	-	-	-	6	4.93
Kell blood group antigen [Homo sapiens]	4557691	-	-	-	-	2	3.28
keratin 2a [Homo sapiens]	4557703	4	6.82	-	-	-	-
neurofibromin; Neurofibromin (neurofibromatosis, type I) [Homo sapiens]	4557793	-	-	-	-	6	2.59
3-oxoacid CoA transferase precursor; Succinyl CoA:3-oxoacid CoA transfe	4557817	-	-	-	-	2	4.62
Propionyl-Coenzyme A carboxylase, alpha polypepti	4557833	-	-	5	6.97	1	1.99
transferrin [Homo sapiens]	4557871	-	-	2	4.82	8	10.03
A Chain A, Structure Of Human Apolactoferrin At 2.0 A Resolution.	4699853	-	-	-	-	3	5.21
DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5; DEAD/H (Asp-Glu-Ala-Asp	4758138	-	-	-	-	6	9.77
FK506-binding protein 5; 51 kDa FK506-binding protein 5; 54 kDa progest	4758384	-	-	-	-	5	12.47
NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q	4758792	-	-	-	-	2	19.35

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
N-myc and STAT interactor; N-myc interactor [Homo sapiens]	4758814	-	-	4	14.66	-	-
PTPL1-associated RhoGAP 1 [Homo sapiens]	4758882	-	-	-	-	8	5.00
phospholipase C, beta 2 [Homo sapiens]	4758938	2	2.12	-	-	-	-
solute carrier family 9 (sodium/hydrogen exchanger), isoform 5 [Homo sa	4759144	-	-	-	-	2	2.34
FK506 binding protein 12-rapamycin associated protein 1; FK506 binding	4826730	-	-	-	-	9	3.18
haptoglobin [Homo sapiens]	4826762	-	-	11	29.56	11	29.56
hydroxysteroid (11-beta) dehydrogenase 1 [Homo sapiens]	5031765	-	-	-	-	2	7.19
endothelial lipase precursor; endothelial cell-derived lipase [Homo sap	5174497	3	5.20	-	-	-	-
nuclear mitotic apparatus protein 1 [Homo sapiens]	5453820	-	-	8	4.14	-	-
pericentriolar material 1 [Homo sapiens]	5453856	-	-	-	-	8	4.79
v-rel reticuloendotheliosis viral oncogene homolog B, nuclear factor of	5730007	3	6.91	-	-	3	4.84
utrophin; dystrophin-related protein [Homo sapien	6005938	-	-	-	-	22	6.35
A Chain A, The Sulfotransferase Domain Of Human Haparin Sulfate N-DeacetylaseN-	6137498	-	-	-	-	5	16.00
CASP8 associated protein 2; FLICE associated huge; human FLASH; FLASH h	6912288	-	-	-	-	11	5.45
zinc finger protein 281; ZNP-99 transcription factor [Homo sapiens]	6912752	-	-	-	-	4	4.92
cell division protein FtsJ [Homo sapiens]	7019377	-	-	-	-	2	9.76
HUMAGCGB protein [Homo sapiens]	7110645	-	-	9	7.81	1	5.47

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
retina and anterior neural fold homeobox; retinal homeobox protein [Hom	7305451	-	-	-	-	3	9.54
JW0088	7435179	-	-	-	-	5	7.43
T12525 hypothetical protein DKFZp434L243.1 - human (fragment)	7512652	-	-	-	-	3	5.20
T00079 hypothetical protein KIAA0465 - human (fragment) s typ	7512993	-	-	-	-	7	4.54
T00383 KIAA0632 protein - human (fragment)	7513104	-	-	-	-	6	8.12
T13163 Rab6 GTPase activating protein, GAPCenA - human Contai	7513292	-	-	-	-	5	4.08
pescadillo homolog 1, containing BRCT domain; pes	7657455	-	-	-	-	6	11.90
DnaJ (Hsp40) homolog, subfamily C, member 8; splicing factor similar to	7657611	-	-	-	-	2	6.82
leucine-tRNA ligase precursor; leucine translase [Homo sapiens]	7661872	-	-	-	-	2	2.55
KIAA0161 gene product [Homo sapiens]	7661956	-	-	-	-	4	15.75
lipin 2 [Homo sapiens]	7662022	-	-	-	-	2	2.79
KIAA0586 gene product [Homo sapiens]	7662190	-	-	-	-	5	3.91
VCL isoform meta-VCL [Homo sapiens]	7669550	-	-	9	4.14	-	-
epithelial protein lost in neoplasm beta; sterol regulatory element bin	7705373	-	-	-	-	2	3.16
beta-ureidopropionase [Homo sapiens]	7706509	-	-	-	-	2	2.08
orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1); a	9257232	-	-	-	-	2	10.45
tuftelin 1 [Homo sapiens]	9910596	-	-	-	-	4	12.82
ATP-binding cassette, sub-family B, member 10 [Homo sapiens]	9961244	4	5.42	3	2.14	4	5.42

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
hemopexin [Homo sapiens]	11321561	-	-	2	3.68	9	15.15
T47182 hypothetical protein DKFZp434M1616.1 - human (fragment)	11360120	-	-	-	-	2	1.23
T46337 hypothetical protein DKFZp434O2413.1 - human (fragment)	11360154	4	4.44	-	-	4	4.44
T42645 hypothetical protein DKFZp566O224.1 - human	11360197	-	-	2	5.99	-	-
RYK receptor-like tyrosine kinase precursor; tyrosylprotein kinase; hy	11863159	-	-	2	3.64	-	-
ZN93_HUMAN Zinc finger protein 93 (Zinc finger protein HTF34)	12643428	-	-	-	-	6	11.69
NCR1_HUMAN Nuclear receptor co-repressor 1 (N-CoR1) (N-CoR)	12643638	-	-	-	-	12	5.41
WAIP_HUMAN Wiskott-Aldrich syndrome protein interacting protein (WASP interac	13124642	-	-	-	-	2	5.17
zinc finger protein 117 (HPF9); Zinc finger protein-117 [Homo sapiens]	13374557	-	-	-	-	2	6.27
integral membrane protein 3; E25 protein [Homo sapiens]	13569885	-	-	-	-	2	8.99
ZO2_HUMAN Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona occl	13634076	-	-	-	-	12	10.25
KF4A_HUMAN Chromosome-associated kinesin KIF4A (Chromokinesin)	13959694	-	-	-	-	7	6.09
allograft inflammatory factor 1, isoform 3; inte	14574568	-	-	-	-	2	8.84
VAC1_HUMAN Vasopressin-activated calcium-mobilizing receptor (VACM-1) (Cullin	14917099	-	-	-	-	11	13.97
keratin 14; cytokeratin 14 [Homo sapiens]	15431310	-	-	6	10.17	4	8.05

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
SPCQ_HUMAN Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3	17368942	6	2.81	-	-	-	-
MOT3_HUMAN Monocarboxylate transporter 3 (MCT 3)	17433295	-	-	-	-	2	3.77
ABG1_HUMAN ATP-binding cassette, sub-family G, member 1 (White protein homolo	17433715	-	-	-	-	3	5.46
chromosome 20 open reading frame 6 [Homo sapiens	18093112	-	-	-	-	12	12.81
A3B2_HUMAN Adapter-related protein complex 3 beta 2 subunit (Beta-adaptin 3B)	18202497	-	-	-	-	5	4.71
enhancer of zeste homolog 1 (Drosophila); enhancer of zeste (Drosophil	19923202	-	-	-	-	5	6.43
chromosome 20 open reading frame 1; chromosome 20 open reading frame 2	20127519	-	-	4	5.35	-	-
PRES_HUMAN Prestin arrier family 21 member 12 (Sodium-i	20139418	-	-	-	-	3	4.41
ACLY_HUMAN ATP-citrate (pro-S-)-lyase (Citrate cleavage enzyme)	20141248	-	-	6	5.63	-	-
NDR2_HUMAN NDRG2 protein (Syld709613 protein)	20141615	-	-	2	6.47	-	-
ZF64_HUMAN Zinc finger protein clone 647	20141866	-	-	-	-	4	10.39
PRTP_HUMAN Lysosomal protective protein precursor (Cathepsin A) (Carboxypepti	20178316	-	-	-	-	2	5.21
NCK2_HUMAN Cytoplasmic protein NCK2 (NCK adaptor protein 2) (SH2/SH3 adaptor	20532395	3	10.21	-	-	8	22.89
similar to hypothetical protein DKFZp586O1822	20883380	-	-	-	-	5	1.99

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
HIK2_HUMAN Homeodomain-interacting protein kinase 2 ent	21431782	-	-	-	-	5	4.01
S100 calcium-binding protein A8; cystic fibrosis antigen; calgranulin	21614544	-	-	-	-	4	24.73
A Chain A, Crystal Structure Of Human Estrogen Sulfotran	21730331	-	-	-	-	6	22.11
NEK1_HUMAN Serine/threonine-protein kinase NEK1 (NimA-related protein kinase	22256934	-	-	15	4.77	4	3.02
BA2B_HUMAN Bromodomain adjacent to zinc finger domain 2B (hWALp4)	22653668	6	2.94	-	-	6	2.94
LMBT_HUMAN Lethal(3)malignant brain tumor-like protein (L(3)mbt-like) (L(3)mb	23396689	-	-	-	-	4	5.70
SG3_HUMAN Secretogranin III precursor (SgIII)	23396839	-	-	-	-	3	5.13
RRB1_HUMAN Ribosome binding protein 1 (Ribosome recepto	23822112	-	-	-	-	10	6.95
trichohyalin [Homo sapiens]	23943906	-	-	9	10.59	17	9.43
TRABID protein [Homo sapiens]	23943914	-	-	7	9.89	-	-
zizimin1 [Homo sapiens]	24308029	-	-	-	-	8	4.88
glutathione S-transferase A3; glutathione S-alkyltransferase A3; gluta	24430144	-	-	-	-	3	14.86
B Chain B, Crystal Structure Of Recombinant Human Fibrinogen Fragment D	24987624	-	-	-	-	4	13.10
A59434 KIAA1501 protein [imported] - human	25535862	-	-	6	8.12	4	5.99
keratin 6 isoform K6e [Homo sapiens]	27465517	-	-	3	5.85	3	5.85
similar to hypothetical protein c316G12.1 (KIAA0	27668068	-	-	-	-	6	5.10

List of human proteins found in CSF Cont'd

Reference	Accession	CM Peptides	CM Peptides % Coverage by AA	EN Peptides	EN Peptides % Coverage by AA	ABM Peptides	ABM Peptides % Coverage by AA
similar to ubiquitin protein ligase E3A, isoform 1; human papilloma vi	27685325	-	-	-	-	7	8.79
SI:dZ256P4.1 (novel protein similar to human prot	27884111	-	-	-	-	6	4.73

APPENDIX IV List of *falciparum* proteins Identified from CSF samples.

Reference	Accession	Cerebral Malaria Peptides	CM %Coverage by AA	EN Peptides	EN %Coverage by AA	ABM Peptides	ABM %Coverage by AA
MAL13P1.144 MAL13P1.144 hypothetical protein 18172617:18175528 reverse MW:56034	MAL13P1.144					9	21.62
MAL13P1.234 MAL13P1.234 hypothetical protein 18926092:18944453 forward MW:696786	MAL13P1.234					117	18.52
MAL13P1.256 MAL13P1.256 phosphatidylinositol transfer protein, putative 19107032:19112954 forward M	MAL13P1.256	24	13.04			32	17.03
MAL7P1.134 MAL7P1.134 hypothetical protein 7707576:7717876 reverse MW:404165	MAL7P1.134					78	20.89
MAL7P1.167 MAL7P1.167 hypothetical protein 7977661:7985982 reverse MW:329471	MAL7P1.167	46	14.97			55	17.71
MAL8P1.103 MAL8P1.103 hypothetical protein, conserved 8695526:8697598 forward MW:64411	MAL8P1.103					13	20.62
MAL8P1.139 MAL8P1.139 hypothetical protein 8369311:8387664 forward MW:705830	MAL8P1.139					118	18.02
MAL8P1.59 MAL8P1.59 hypothetical protein 9028411:9029920 reverse MW:43129	MAL8P1.59					11	25.07
PF07_0035 PF07_0035 cg1 protein 7079984:7083730 forward MW:146630	PF07_0035					39	19.23
PF07_0071 PF07_0071 queuine tRNA-ribosyltransferase; putative 7423644:7425899 forward MW:88035	PF07_0071					12	15.31
PF07_0086 PF07_0086 hypothetical protein 7575910:7586438 forward MW:416638	PF07_0086					77	22.40
PF08_0080 PF08_0080 hypothetical protein 8784777:8786324 reverse MW:61577	PF08_0080					14	23.30

List of falciparum proteins found in CSF Cont'd

Reference	Accession	Cerebral Malaria Peptides	CM %Coverage by AA	EN Peptides	EN %Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF10_0115 PF10_0115 QF122 antigen 11534791:11538210 forward MW:131627	PF10_0115					31	26.87
PF10_0267 PF10_0267 hypothetical protein 12217394:12218956 forward MW:61468	PF10_0267	10	8.17			25	38.65
PF11_0115 PF11_0115 hypothetical protein 13198955:13200694 reverse MW:69018	PF11_0115					14	21.76
PF11_0129 PF11_0129 hypothetical protein 13238995:13241723 reverse MW:102084	PF11_0129					15	16.45
PF11_0374 PF11_0374 hypothetical protein 14189790:14193247 forward MW:128546	PF11_0374	20	15.94			33	25.14
PF11_0392 PF11_0392 hypothetical protein 14254640:14263189 forward MW:335925	PF11_0392					60	18.88
PF11_0468 PF11_0468 hypothetical protein 14580465:14581619 forward MW:46253	PF11_0468					7	17.71
PF13_0027 PF13_0027 hypothetical protein 17348476:17350791 forward MW:92277	PF13_0027					13	12.97
PF13_0078 PF13_0078 hypothetical protein 17650234:17655544 reverse MW:209078	PF13_0078					28	10.02
PF13_0091 PF13_0091 hypothetical protein 17763016:17765736 forward MW:106123	PF13_0091					19	18.21
PF13_0292 PF13_0292 hypothetical protein 19264254:19267307 reverse MW:121356	PF13_0292					24	19.57
PF14_0002 PF14_0002 rifin 19979677:19980609 reverse MW:34414	PF14_0002					6	19.03
PF14_0081 PF14_0081 DNA repair helicase, putative 20286443:20289925 reverse MW:135547	PF14_0081					25	21.81
PF14_0112 PF14_0112 POM1, putative 20433869:20439919 reverse MW:235822	PF14_0112					36	18.20

List of falciparum proteins found in CSF Cont'd

Reference	Accession	Cerebral Malaria Peptides	CM %Coverage by AA	EN Peptides	EN %Coverage by AA	ABM Peptides	ABM %Coverage by AA
PF14_0169 PF14_0169 hypothetical protein 20662918:20663739 forward MW:32338	PF14_0169					9	35.16
PF14_0175 PF14_0175 hypothetical protein 20703161:20717149 reverse MW:548876	PF14_0175					80	15.85
PF14_0320 PF14_0320 hypothetical protein 21332413:21336969 forward MW:186675	PF14_0320					56	36.63
PF14_0326 PF14_0326 hypothetical protein 21350759:21375076 reverse MW:969736	PF14_0326					164	19.08
PF14_0334 PF14_0334 NAD(P)H-dependent glutamate synthase, putative 21394312:21403632 reverse MW:353	PF14_0334					53	17.00
PF14_0364 PF14_0364 cleavage and polyadenylation specifity factor protein, putative 21529938:215325	PF14_0364	9	10.39			14	16.55
PF14_0372 PF14_0372 hypothetical protein 21566916:21572918 forward MW:221690	PF14_0372					36	14.69
PF14_0392 PF14_0392 Ser/Thr protein kinase, putative 21654202:21661568 forward MW:263797	PF14_0392					57	22.12
PF14_0419 PF14_0419 hypothetical protein 21774372:21797737 reverse MW:857945	PF14_0419					153	21.42
PF14_0509 PF14_0509 hypothetical protein 22171775:22177112 reverse MW:208230	PF14_0509					32	18.43
PF14_0594 PF14_0594 hypothetical protein 22503835:22513803 reverse MW:395683	PF14_0594					69	20.32
PF14_0626 PF14_0626 dynein beta chain, putative 22634424:22655109 forward MW:771768	PF14_0626					115	17.92
PF14_0712 PF14_0712 hypothetical protein 23016532:23027462 forward MW:353912	PF14_0712					75	25.01

List of falciparum proteins found in CSF Cont'd

Reference	Accession	Cerebral Malaria Peptides	CM %Coverage by AA	EN Peptides	EN %Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFA0180w PFA0180w hypothetical protein 161365:166464 forward MW:175729	PFA0180w					30	20.38
PFA0290w PFA0290w DNA binding protein, putative 253399:255169 forward MW:54941	PFA0290w					7	12.15
PFB0755w PFB0755w hypothetical protein 1321778:1326994 forward MW:197779	PFB0755w			38	21.42		
PFB0765w PFB0765w hypothetical protein 1333602:1339074 forward MW:167009	PFB0765w	42	27.77				
PFB0770c PFB0770c hypothetical protein 1339484:1342606 reverse MW:122455	PFB0770c					16	17.21
PFC0415c PFC0415c hypothetical protein 2001302:2006741 reverse MW:203955	PFC0415c					41	21.10
PFC1015c PFC1015c hypothetical protein 2542411:2549433 reverse MW:280921	PFC1015c					57	21.32
PFD0200c PFD0200c hypothetical protein 2881723:2888678 reverse MW:271093	PFD0200c					43	18.01
PFD0535w PFD0535w hypothetical protein 3143210:3147106 forward MW:150271	PFD0535w	18	13.48				
PFD1160w PFD1160w hypothetical protein 3754732:3762032 forward MW:286418	PFD1160w			78	26.55		
PFE0240c PFE0240c hypothetical protein 4042406:4059325 reverse MW:672034	PFE0240c					128	20.75
PFE0270c PFE0270c DNA repair protein, putative 4094049:4098473 reverse MW:156441	PFE0270c					38	26.22
PFE1095w PFE1095w hypothetical protein 4743247:4749317 forward MW:213324	PFE1095w					45	24.59
PFF0125c PFF0125c hypothetical protein 5305179:5309806 reverse MW:153721	PFF0125c					24	18.27

List of falciparum proteins found in CSF Cont'd

Reference	Accession	Cerebral Malaria Peptides	CM %Coverage by AA	EN Peptides	EN %Coverage by AA	ABM Peptides	ABM %Coverage by AA
PFF1440w PFF1440w SET-domain protein, putative 6420088:6441070 forward MW:797045	PFF1440w					147	18.04
PFF1470c PFF1470c DNA polymerase epsilon, catalytic subunit a, putative 6458546:6467744 reverse MW:	PFF1470c					59	18.85
PFI0175w PFI0175w hypothetical protein 9706162:9710556 forward MW:88672	PFI0175w					27	35.98
PFI0260c PFI0260c hypothetical protein 9789022:9807378 reverse MW:720597	PFI0260c					104	15.82
PFI0665w PFI0665w hypothetical protein 10122388:10126083 forward MW:147869	PFI0665w					23	20.47
PFI1120c PFI1120c hypothetical protein 10459953:10473033 reverse MW:509752	PFI1120c			79	18.78		
PFI1580c PFI1580c DHHC-type zinc finger protein, putative 10831930:10834305 reverse MW:62145	PFI1580c					13	24.02
PFL0410w 2277.t00083 hypothetical protein 15181933:15192864 forward MW:418854	PFL0410w					73	18.07
PFL1155w 2277.t00231 GTP cyclohydrolase I 15779749:15780918 forward MW:45877	PFL1155w					9	21.34
PFL1330c 2277.t00266 hypothetical protein 15919401:15926469 reverse MW:273651	PFL1330c					55	23.06
PFL1715w 2277.t00343 hypothetical protein 16286374:16288314 forward MW:75052	PFL1715w					16	24.30
PFL2375c 2277.t00475 cutA, putative 16822736:16823841 reverse MW:18685	PFL2375c					8	45.28
PFL2505c 2277.t00499 hypothetical protein 16919999:16927858 reverse MW:263159	PFL2505c					58	23.07

APPENDIX V List of human proteins found in both plasma and CSF

	Reference	Accession	Plasma	CSF
1	B34611 3',5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35 alpha chain - human	105117	CM, EN, ABM	CM, ABM
2	sperm activating protein subunit I, apolipoprotein A1, SPAP subunit I [hum	235865	EN	ABM
3	K2C1_HUMAN Keratin, type II cytoskeletal 1 (Cytokeratin 1 (K1 (CK 1 (67 kDa	1346343	CM, ABM	CM, EN, ABM
4	A1AT_HUMAN Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor (Alpha-1	1703025	CM, EN, ABM	EN, ABM
5	I38344 titin, cardiac muscle [validated] - human iciency viru	2136280	ABM	ABM
6	S69339 Ig heavy chain V region precursor - human ain-binding	2146957	CM, EN, ABM	EN, ABM
7	A Chain A, Crystal Structure Of Human Apolipoprotein A-I	2914175	EN	ABM
8	C Chain C, Human Serum Transferrin, Recombinant N-Terminal Lobe, Apo Form	4389232	CM, EN, ABM	CM, EN, ABM
9	Human Serum Albumin In A Complex With Myristic Acid And Tri-Iodobenzoic Acid	4389275	CM, EN, ABM	CM, EN, ABM
10	guanylate cyclase 2F; RetGC-2; guanylate cyclase 2D-like, membrane (ret	4504219	CM, ABM	CM
11	delta globin [Homo sapiens]	4504351	EN	CM, EN, ABM
12	orosomucoid 2; alpha-1-acid glycoprotein, type 2 [Homo sapiens]	4505529	EN	ABM
13	secretogranin II precursor; Chromogranin C (secretogranin II ; secreton	4506801	ABM	CM
14	thyrotrophic embryonic factor; Thyrotroph embryonic factor [Homo sapien	4507431	ABM	ABM
15	zinc finger protein 91 (HPF7, HTF10 [Homo sapiens]	4508041	EN, ABM	ABM
16	alpha 2 macroglobulin precursor [Homo sapiens]	4557225	CM, EN, ABM	ABM
17	apolipoprotein A-I precursor [Homo sapiens]	4557321	EN	ABM
18	Propionyl-Coenzyme A carboxylase, alpha polypeptide precursor [Homo sap	4557833	CM, EN	EN, ABM
19	transferrin [Homo sapiens]	4557871	CM, EN, ABM	EN, ABM
20	nuclear mitotic apparatus protein 1 [Homo sapiens	5453820	CM, ABM	EN
21	v-rel reticuloendotheliosis viral oncogene homolog B, nuclear factor of	5730007	CM, ABM	CM, ABM

Human proteins in both plasma and CSF cont'd

	Reference	Accession	Plasma	CSF
22	CASP8 associated protein 2; FLICE associated huge; human FLASH	6912288	CM, EN, ABM	ABM
23	zinc finger protein 281; ZNP-99 transcription factor [Homo sapiens]	6912752	EN	ABM
24	DnaJ (Hsp40 homolog, subfamily C, member 8; splicing factor similar to	7657611	CM, ABM	ABM
25	epithelial protein lost in neoplasm beta; sterol regulatory element bin	7705373	EN, ABM	ABM
26	beta-ureidopropionase [Homo sapiens]	7706509	ABM	ABM
27	orosomucoid 1 precursor; Orosomucoid-1 (alpha-1-acid glycoprotein-1 ; a	9257232	CM	ABM
28	ATP-binding cassette, sub-family B, member 10 [Homo sapiens]	9961244	CM, EN, ABM	CM, EN, ABM
29	T46337 hypothetical protein DKFZp434O2413.1 - human (fragment	11360154	ABM	CM, ABM
30	ZN93_HUMAN Zinc finger protein 93 (Zinc finger protein HTF34	12643428	EN	ABM
31	zinc finger protein 117 (HPF9 ; Zinc finger protein-117 [Homo sapiens]	13374557	EN	ABM
32	VAC1_HUMAN Vasopressin-activated calcium-mobilizing receptor (VACM-1 (Cullin	14917099	ABM	ABM
33	SPCQ_HUMAN Spectrin beta chain, brain 3 (Spectrin, non-	17368942	CM, EN, ABM	CM
34	MOT3_HUMAN Monocarboxylate transporter 3 (MCT 3	17433295	ABM	ABM
35	enhancer of zeste homolog 1 (Drosophila ; enhancer of zeste (Drosophil	19923202	EN	ABM
36	PRES_HUMAN Prestin	20139418	ABM	ABM
37	NDR2_HUMAN NDRG2 protein (Syld709613 protein	20141615	CM, EN, ABM	EN
38	NCK2_HUMAN Cytoplasmic protein NCK2 (NCK adaptor protei	20532395	ABM	CM, ABM
39	BA2B_HUMAN Bromodomain adjacent to zinc finger domain 2B (hWALp4	22653668	ABM	CM, ABM
40	RRB1_HUMAN Ribosome binding protein 1 (Ribosome receptor protein (180 kDa ri	23822112	ABM	ABM
41	similar to ubiquitin protein ligase E3A, isoform 1; human papilloma vi	27685325	ABM	ABM

APPENDIX VI List of *falciparum* proteins found in both plasma and CSF

	Reference	Accession	Plasma	CSF
1	MAL13P1.234 MAL13P1.234 hypothetical protein 18926092:18944453 forward MW:696786	MAL13P1.234	CM, ABM	ABM
2	MAL13P1.256 MAL13P1.256 phosphatidylinositol transfer protein, putative 19107032:19112954 forward M	MAL13P1.256	CM, ABM, EN	CM, ABM
3	MAL7P1.167 MAL7P1.167 hypothetical protein, conserved 7977659:7985980 reverse MW:329471	MAL7P1.167	CM, ABM, EN	CM, ABM
4	PF08_0080 PF08_0080 hypothetical protein 8784777:8786324 reverse MW:61577	PF08_0080	CM	ABM
5	PF10_0267 PF10_0267 hypothetical protein 12217392:12218954 forward MW:61468	PF10_0267	CM, ABM, EN	CM, ABM
6	PF11_0129 PF11_0129 hypothetical protein 13238995:13241723 reverse MW:102084	PF11_0129	CM, ABM, EN	ABM
7	PF11_0374 PF11_0374 hypothetical protein 14189790:14193247 forward MW:128546	PF11_0374	CM, ABM, EN	CM, ABM
8	PF11_0392 PF11_0392 hypothetical protein 14254640:14263189 forward MW:335925	PF11_0392	ABM	ABM
9	PF11_0468 PF11_0468 hypothetical protein 14580463:14581617 forward MW:46253	PF11_0468	EN, ABM	ABM
10	PF13_0078 PF13_0078 hypothetical protein 17650234:17655544 reverse MW:209078	PF13_0078	ABM	ABM
11	PF14_0081 PF14_0081 DNA repair helicase, putative 20286443:20289925 reverse MW:135547	PF14_0081	CM, ABM, EN	ABM
12	PF14_0112 PF14_0112 POM1, putative 20433869:20439919 reverse MW:235822	PF14_0112	CM, ABM	ABM
13	PF14_0175 PF14_0175 hypothetical protein 20703161:20717149 reverse MW:548876	PF14_0175	ABM	ABM
14	PF14_0326 PF14_0326 hypothetical protein 21350759:21375076 reverse MW:969736	PF14_0326	CM, ABM	ABM
15	PF14_0334 PF14_0334 NAD(P)H-dependent glutamate synthase, putative 21394312:21403632 reverse MW:353	PF14_0334	CM	ABM
16	PF14_0364 PF14_0364 cleavage and polyadenylation specifity factor protein, putative 21529938:215325	PF14_0364	CM	CM, ABM
17	PF14_0372 PF14_0372 hypothetical protein 21566916:21572918 forward MW:221690	PF14_0372	CM, ABM, EN	ABM
18	PF14_0392 PF14_0392 Ser/Thr protein kinase, putative 21654202:21661568 forward MW:263797	PF14_0392	ABM	ABM
19	PF14_0419 PF14_0419 hypothetical protein 21774372:21797737 reverse MW:857945	PF14_0419	CM, ABM	ABM
20	PF14_0509 PF14_0509 hypothetical protein 22171773:22177110 reverse MW:208230	PF14_0509	EN	ABM
21	PF14_0626 PF14_0626 dynein beta chain, putative 22634424:22655109 forward MW:771768	PF14_0626	CM, ABM	ABM

Falcjparum proteins in both plasma and CSF cont'd

	Reference	Accession	Plasma	CSF
22	PFA0180w PFA0180w hypothetical protein 161365:166464 forward MW:175729	PFA0180w	CM, ABM	ABM
23	PFA0290w PFA0290w DNA binding protein, putative 253399:255169 forward MW:54941	PFA0290w	EN	ABM
24	PFD0200c PFD0200c hypothetical protein 2881723:2888678 reverse MW:271093	PFD0200c	ABM	ABM
25	PFD0535w PFD0535w hypothetical protein 3143210:3147106 forward MW:150271	PFD0535w	CM, ABM	CM
26	PFD1160w PFD1160w hypothetical protein 3754732:3762032 forward MW:286418	PFD1160w	CM	EN
27	PFE0240c PFE0240c hypothetical protein 4042406:4059325 reverse MW:672034	PFE0240c	CM	ABM
28	PFF1440w PFF1440w SET-domain protein, putative 6420088:6441070 forward MW:797045	PFF1440w	CM, ABM	ABM
29	PFF1470c PFF1470c DNA polymerase epsilon, catalytic subunit a, putative 6458546:6467744 reverse MW:	PFF1470c	ABM	ABM
30	PFI0260c PFI0260c hypothetical protein 9789022:9807378 reverse MW:720597	PFI0260c	CM, ABM	ABM
31	PFI0665w PFI0665w hypothetical protein 10122388:10126083 forward MW:147869	PFI0665w	CM, ABM, EN	ABM
32	PFL0410w 2277.t00083 hypothetical protein 15181933:15192864 forward MW:418854	PFL0410w	CM	ABM
33	PFL1330c 2277.t00266 hypothetical protein 15919401:15926469 reverse MW:273651	PFL1330c	CM, ABM, EN	ABM
34	PFL2375c 2277.t00475 cutA, putative 16822736:16823841 reverse MW:18685	PFL2375c	ABM	ABM

APPENDIX VII List of human proteins found in plasma with Gene Ontology (GO).

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Rod cGMP-specific 3',5'-cyclic phosphodiesterase subunit alpha (EC 3.1.4.35) (GMP-PDE alpha) (PDE V-B1)	105117	0016787 : hydrolase activity; 0003824 : catalytic activity	0016020 : membrane	0008150 : biological_process; 0007165 : signal transduction; 0050896 : response to stimulus
Apolipoprotein A-IV precursor (Apo-AIV) (ApoA-IV)	114006	0005215 : transporter activity; 0016209 : antioxidant activity; 0043167 : ion binding; 0008289 : lipid binding; 0005515 : protein binding	0005576 : extracellular region	0019538 : protein metabolism; 0006800 : oxygen and reactive oxygen species metabolism; 0050789 : regulation of biological process; 0050896 : response to stimulus; 0006629 : lipid metabolism; 0006869 : lipid transport; 0006810 : transport; 0008150 : biological_process
Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD)	120731	0016491 : oxidoreductase activity	0005737 : cytoplasm	0005975 : carbohydrate metabolism; 0006066 : alcohol metabolism; 0006766 : vitamin metabolism; 0009117 : nucleotide metabolism; 0051186 : cofactor metabolism
Prostaglandin G/H synthase 1 precursor (EC 1.14.99.1) (Cyclooxygenase- 1) (COX-1) (Prostaglandin-endoperoxide synthase 1) (Prostaglandin H2 synthase 1) (PGH synthase 1) (PGHS-1) (PHS 1)	129899	0016491 : oxidoreductase activity; 0043167 : ion binding; 0016209 : antioxidant activity	0016020 : membrane; 0005737 : cytoplasm; 0000267 : cell fraction; 0005634 : nucleus	0006629 : lipid metabolism; 0009987 : cellular process; 0006082 : organic acid metabolism; 0008150 : biological_process
Spectrin beta chain, erythrocyte (Beta-I spectrin)	134798	0005515 : protein binding; 0005198 : structural molecule activity	0016020 : membrane; 0043234 : protein complex; 0005856 : cytoskeleton; 0005737 : cytoplasm; 0000267 : cell fraction	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Sperm activating protein subunit I, apolipoprotein A1, SPAP subunit I (Fragment)	235865			
14-3-3 protein eta (Protein AS1)	481365	0005515 : protein binding; 0030528 : transcription regulator activity; 0030234 : enzyme regulator activity	0005737 : cytoplasm	0006629 : lipid metabolism; 0042445 : hormone metabolism; 0006350 : transcription; 0050789 : regulation of biological process; 0000902 : cellular morphogenesis; 0007275 : development; 0007049 : cell cycle; 0009987 : cellular process; 0007165 : signal transduction; 0008150 : biological_process; 0008219 : cell death; 0015031 : protein transport; 0046907 : intracellular transport; 0016043 : cell organization and biogenesis
Non-POU domain-containing octamer-binding protein (NonO protein) (54 kDa nuclear RNA- and DNA-binding protein) (p54(nrb)) (p54nrb) (55 kDa nuclear protein) (NMT55) (DNA-binding p52/p100 complex, 52 kDa subunit)	543010	0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0005515 : protein binding	0005634 : nucleus	0006310 : DNA recombination; 0006396 : RNA processing; 0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process; 0050896 : response to stimulus; 0006281 : DNA repair

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Centromere protein F (Kinetochore protein CENP-F) (Mitosin) (AH antigen)	1345731	0005515 : protein binding; 0005488 : binding	0005737 : cytoplasm; 0005694 : chromosome; 0043234 : protein complex; 0005819 : spindle; 0005634 : nucleus; 0016020 : membrane	0007049 : cell cycle; 0007059 : chromosome segregation; 0050789 : regulation of biological process; 0007275 : development; 0051301 : cell division; 0016043 : cell organization and biogenesis; 0051179 : localization; 0050896 : response to stimulus; 0006350 : transcription
Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1) (K1) (67 kDa cytokeratin) (Hair alpha protein)	1346343	0030246 : carbohydrate binding; 0004871 : signal transducer activity; 0005198 : structural molecule activity; 0005515 : protein binding	0016020 : membrane; 0005856 : cytoskeleton	0050789 : regulation of biological process; 0050817 : coagulation; 0050896 : response to stimulus; 0006800 : oxygen and reactive oxygen species metabolism; 0007275 : development; 0019538 : protein metabolism
Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1- antiproteinase)	1703025	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus
Dihydrolipoyllysine-residue succinyltransferase component of 2- oxoglutarate dehydrogenase complex, mitochondrial precursor (EC 2.3.1.61) (Dihydrolipoamide succinyltransferase component of 2- oxoglutarate dehydrogenase complex) (E2) (E2K)	1709442	0016740 : transferase activity; 0031406 : carboxylic acid binding; 0048037 : cofactor binding	0005739 : mitochondrion; 0005622 : intracellular; 0043234 : protein complex	0008152 : metabolism; 0006091 : generation of precursor metabolites and energy; 0005975 : carbohydrate metabolism; 0051186 : cofactor metabolism
DNA2-like helicase (EC 3.6.1.-) (DNA replication ATP-dependent helicase-like homolog)	2506893	0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0016787 : hydrolase activity; 0003824 : catalytic activity		0006260 : DNA replication

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Cell division control protein 42 homolog precursor (G25K GTP-binding protein)	2624582	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0005515 : protein binding	0016020 : membrane; 0005737 : cytoplasm; 0005622 : intracellular; 0042995 : cell projection	0016043 : cell organization and biogenesis; 0006461 : protein complex assembly; 0050789 : regulation of biological process; 0007165 : signal transduction; 0000902 : cellular morphogenesis; 0007275 : development; 0009987 : cellular process
Cytoskeleton-associated protein 5 (Colonic and hepatic tumor over- expressed protein) (Ch-TOG protein)	3121951	0005488 : binding; 0005515 : protein binding	0005737 : cytoplasm; 0005856 : cytoskeleton; 0043234 : protein complex; 0005813 : centrosome	0015931 : nucleobase\, nucleoside\, nucleotide and nucleic acid transport; 0006996 : organelle organization and biogenesis; 0016043 : cell organization and biogenesis; 0007049 : cell cycle
DNA polymerase zeta catalytic subunit (EC 2.7.7.7) (hREV3)	3913527	0016740 : transferase activity; 0003676 : nucleic acid binding; 0016787 : hydrolase activity; 0000166 : nucleotide binding; 0043167 : ion binding	0005634 : nucleus; 0042575 : DNA polymerase complex; 0043229 : intracellular organelle	0006260 : DNA replication; 0050896 : response to stimulus; 0006281 : DNA repair
Serum albumin (Cell growth inhibiting protein 42)	4502027	0005215 : transporter activity	0005576 : extracellular region	0006810 : transport
Aminopeptidase N (EC 3.4.11.2) (hAPN) (Alanyl aminopeptidase) (Microsomal aminopeptidase) (Aminopeptidase M) (gp150) (Myeloid plasma membrane glycoprotein CD13) (CD13 antigen)	4502095	0016787 : hydrolase activity; 0004871 : signal transducer activity; 0043167 : ion binding	0016020 : membrane; 0005737 : cytoplasm	0007275 : development; 0009987 : cellular process; 0019538 : protein metabolism

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Calcium/calmodulin-dependent protein kinase type IV (EC 2.7.11.17) (CAM kinase-GR) (CaMK IV)	4502557	0016740 : transferase activity; 0030528 : transcription regulator activity; 0000166 : nucleotide binding; 0005515 : protein binding; 0043167 : ion binding	0005737 : cytoplasm; 0005634 : nucleus	0007267 : cell-cell signaling; 0007165 : signal transduction; 0046907 : intracellular transport; 0006464 : protein modification; 0006793 : phosphorus metabolism
Eukaryotic translation initiation factor 4E (eIF4E) (eIF-4E) (mRNA cap-binding protein) (eIF-4F 25 kDa subunit)	4503535	0003676 : nucleic acid binding; 0045182 : translation regulator activity; 0005515 : protein binding	0005737 : cytoplasm; 0043234 : protein complex	0006412 : protein biosynthesis; 0016043 : cell organization and biogenesis; 0050789 : regulation of biological process
Coagulation factor VIII precursor (Procoagulant component) (Antihemophilic factor) (AHF) [Contains: Factor VIIIa heavy chain, 200 kDa isoform; Factor VIIIa heavy chain, 92 kDa isoform; Factor VIII B chain; Factor VIIIa light chain]	4503647	0016491 : oxidoreductase activity; 0043167 : ion binding	0005576 : extracellular region	0050817 : coagulation; 0050896 : response to stimulus; 0007155 : cell adhesion
Coagulation factor VIII, procoagulant component (Hemophilia A)	4503647	0016491 : oxidoreductase activity; 0043167 : ion binding		0050817 : coagulation; 0050896 : response to stimulus; 0007155 : cell adhesion
Very-long-chain acyl-CoA synthetase (EC 6.2.1.-) (VLCS) (Very-long- chain-fatty-acid-CoA ligase) (VLACS) (THCA-CoA ligase) (Fatty-acid- coenzyme A ligase, very long-chain 1) (Long-chain-fatty-acid--CoA ligase) (EC 6.2.1.3) (Fatty acid transport protein 2) (FATP-2) (Solute carrier family 27 member 2)	4503653	0003824 : catalytic activity; 0000166 : nucleotide binding; 0016874 : ligase activity	0016020 : membrane; 0005739 : mitochondrion; 0005783 : endoplasmic reticulum; 0042579 : microbody	0008152 : metabolism; 0006629 : lipid metabolism; 0006082 : organic acid metabolism

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Trifunctional purine biosynthetic protein adenosine-3 [Includes: Phosphoribosylamine--glycine ligase (EC 6.3.4.13) (GARS) (Glycinamide ribonucleotide synthetase) (Phosphoribosylglycinamide synthetase); Phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1) (AIRS) (Phosphoribosyl-aminoimidazole synthetase) (AIR synthase); Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR transformylase) (5'-phosphoribosylglycinamide transformylase)]	4503915	0016874 : ligase activity; 0003824 : catalytic activity; 0000166 : nucleotide binding; 0016740 : transferase activity	0005737 : cytoplasm	0009058 : biosynthesis; 0006412 : protein biosynthesis; 0009112 : nucleobase metabolism; 0042440 : pigment metabolism; 0009117 : nucleotide metabolism
Delta-globin chain (Hemoglobin delta)	4504351	0005488 : binding; 0019825 : oxygen binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport
Hemoglobin subunit delta (Hemoglobin delta chain) (Delta-globin)	4504351	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport
NFkappaB essential modulator (IKBKG protein)	4504631			
Down syndrome critical region protein 2 (Leucine-rich protein C21- LRP)	4505023		0016020 : membrane; 0000267 : cell fraction	
DSCR2 protein (Proteasome assembling chaperone-1)	4505023		0043234 : protein complex; 0005737 : cytoplasm	
Alpha-1-acid glycoprotein 2 precursor (AGP 2) (Orosomucoid-2) (OMD 2)	4505529	0005488 : binding	0005576 : extracellular region	0050896 : response to stimulus
Orosomucoid 2	4505529	0005488 : binding		

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Proteasome subunit alpha type 2 (EC 3.4.25.1) (Proteasome component C3) (Macropain subunit C3) (Multicatalytic endopeptidase complex subunit C3)	4506181	0016787 : hydrolase activity	0000502 : proteasome complex (sensu Eukaryota); 0043234 : protein complex; 0005737 : cytoplasm	0019538 : protein metabolism
Proteasome subunit alpha type (EC 3.4.25.1) (Fragment)	4506181	0016787 : hydrolase activity	0043234 : protein complex; 0005737 : cytoplasm; 0000502 : proteasome complex (sensu Eukaryota)	0019538 : protein metabolism
60S ribosomal protein L24 (Ribosomal protein L30)	4506619	0005198 : structural molecule activity; 0005515 : protein binding; 0003676 : nucleic acid binding	0005840 : ribosome; 0043234 : protein complex; 0005622 : intracellular; 0030529 : ribonucleoprotein complex	0006412 : protein biosynthesis
60S ribosomal protein L32	4506635	0005198 : structural molecule activity	0005622 : intracellular; 0005840 : ribosome; 0030529 : ribonucleoprotein complex	0006412 : protein biosynthesis
CAP-Gly domain-containing linker protein 1 (Restin) (Cytoplasmic linker protein 170 alpha-2) (CLIP-170) (Reed-Sternberg intermediate filament-associated protein) (Cytoplasmic linker protein 1)	4506751	0005515 : protein binding; 0003676 : nucleic acid binding; 0043167 : ion binding	0005694 : chromosome; 0043234 : protein complex; 0005856 : cytoskeleton; 0005768 : endosome	0007049 : cell cycle

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Amiloride-sensitive sodium channel subunit alpha (Epithelial Na(+) channel subunit alpha) (Alpha ENaC) (Nonvoltage-gated sodium channel 1 subunit alpha) (SCNEA) (Alpha NaCH)	4506815	0043167 : ion binding; 0005215 : transporter activity; 0005515 : protein binding	0016020 : membrane; 0000267 : cell fraction	0006811 : ion transport; 0006810 : transport; 0050896 : response to stimulus; 0008150 : biological_process
Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase)	4507155	0016491 : oxidoreductase activity; 0043167 : ion binding		0008150 : biological_process; 0005975 : carbohydrate metabolism; 0006066 : alcohol metabolism
Thyrotroph embryonic factor	4507431	0030528 : transcription regulator activity; 0003676 : nucleic acid binding; 0005515 : protein binding	0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process; 0048511 : rhythmic process
Transforming growth factor-beta-induced protein ig-h3 precursor (Beta ig-h3) (Kerato-epithelin) (RGD-containing collagen-associated protein) (RGD-CAP)	4507467	0005515 : protein binding	0005576 : extracellular region; 0031012 : extracellular matrix	0008150 : biological_process; 0007155 : cell adhesion; 0050789 : regulation of biological process; 0050896 : response to stimulus; 0008283 : cell proliferation
Transforming growth factor, beta-induced, 68kDa	4507467			0007155 : cell adhesion
Cylicin-2 (Cylicin II) (Multiple-band polypeptide II)	4557509	0005198 : structural molecule activity	0005856 : cytoskeleton	0009987 : cellular process; 0007275 : development; 0008150 : biological_process
Fatty acid-binding protein, epidermal (E-FABP) (Psoriasis-associated fatty acid-binding protein homolog) (PA-FABP)	4557581	0005488 : binding; 0005515 : protein binding; 0008289 : lipid binding	0005737 : cytoplasm	0007275 : development; 0006629 : lipid metabolism; 0006810 : transport

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Transferrin	4557871	0043167 : ion binding	0005576 : extracellular region; 0016020 : membrane; 0045177 : apical part of cell; 0016023 : cytoplasmic membrane-bound vesicle	0019725 : cell homeostasis; 0006811 : ion transport
Spliceosome RNA helicase BAT1 (EC 3.6.1.-) (DEAD box protein UAP56) (56 kDa U2AF65-associated protein) (ATP-dependent RNA helicase p47) (HLA-B-associated transcript-1)	4758112	0016787 : hydrolase activity; 0005515 : protein binding; 0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity	0005681 : spliceosome complex; 0005634 : nucleus	0006396 : RNA processing; 0015931 : nucleobase\, nucleoside\, nucleotide and nucleic acid transport; 0046907 : intracellular transport
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (NADH-ubiquinone oxidoreductase 30 kDa subunit) (Complex I-30kD) (CI-30kD)	4758788	0016491 : oxidoreductase activity	0005739 : mitochondrion; 0000267 : cell fraction	0006091 : generation of precursor metabolites and energy; 0006793 : phosphorus metabolism
Ras-related protein Rab-28 (Rab-26)	4758994	0016787 : hydrolase activity; 0000166 : nucleotide binding	0016020 : membrane	0007165 : signal transduction
Transcription factor SOX-14	4759162	0003676 : nucleic acid binding; 0030528 : transcription regulator activity	0005694 : chromosome; 0005634 : nucleus	0007275 : development; 0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Vesicle-associated membrane protein-associated protein B/C (VAMP- associated protein B/C) (VAMP-B/VAMP-C) (VAP-B/VAP-C)	4759302	0005198 : structural molecule activity	0016020 : membrane	0006461 : protein complex assembly
VAMP (Vesicle-associated membrane protein)-associated protein B and C	4759302	0005198 : structural molecule activity		

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
CDNA FLJ45319 fis, clone BRHIP3005801, highly similar to Homo sapiens VAMP (vesicle-associated membrane protein)-associated protein B and C (VAPB)	4759302			
Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] 2 (EC 2.6.1.16) (Hexosephosphate aminotransferase 2) (D-fructose-6- phosphate amidotransferase 2) (GFAT 2) (GFAT2)	4826742	0016740 : transferase activity; 0030246 : carbohydrate binding	0005737 : cytoplasm	0005975 : carbohydrate metabolism; 0006066 : alcohol metabolism; 0008152 : metabolism; 0006520 : amino acid metabolism; 0006091 : generation of precursor metabolites and energy
Matrix metalloproteinase-14 precursor (EC 3.4.24.80) (MMP-14) (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1) (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MT1MMP) (MMP- X1)	4826834	0016787 : hydrolase activity; 0043167 : ion binding; 0005515 : protein binding	0016020 : membrane; 0005576 : extracellular region; 0031012 : extracellular matrix	0005975 : carbohydrate metabolism; 0019538 : protein metabolism
Matrix metalloproteinase 1 (Fragment)	4826834			
Ubiquitin-conjugating enzyme E2-25 kDa (EC 6.3.2.19) (Ubiquitin- protein ligase) (Ubiquitin carrier protein) (E2(25K)) (Huntingtin-interacting protein 2) (HIP-2)	4885417	0016874 : ligase activity		0019538 : protein metabolism; 0006464 : protein modification
Proto-oncogene protein Wnt-1 precursor	4885655	0004871 : signal transducer activity; 0005515 : protein binding	0000267 : cell fraction; 0005576 : extracellular region	0007275 : development; 0009987 : cellular process; 0007165 : signal transduction; 0000902 : cellular morphogenesis; 0006928 : cell motility; 0008150 : biological_process
Wingless-type MMTV integration site family, member 1	4885655	0004871 : signal transducer activity	0005576 : extracellular region	0007165 : signal transduction; 0007275 : development
Microtubule-associated protein 1B	5174525			

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1]	5174525	0005198 : structural molecule activity; 0005515 : protein binding	0005856 : cytoskeleton; 0043234 : protein complex	0000902 : cellular morphogenesis; 0007275 : development; 0016043 : cell organization and biogenesis
Mitochondrial intermediate peptidase, mitochondrial precursor (EC 3.4.24.59) (MIP)	5174567	0016787 : hydrolase activity; 0043167 : ion binding	0005739 : mitochondrion	0019538 : protein metabolism; 0006464 : protein modification; 0015031 : protein transport; 0046907 : intracellular transport
MSTP039	5454156	0016787 : hydrolase activity		0019538 : protein metabolism
Histone acetyltransferase MYST3 (EC 2.3.1.48) (EC 2.3.1.-) (MYST protein 3) (MOZ, YBF2/SAS3, SAS2 and TIP60 protein 3) (Runt-related transcription factor-binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc finger protein 220)	5803098	0003676 : nucleic acid binding; 0005515 : protein binding; 0016740 : transferase activity; 0043167 : ion binding	0005694 : chromosome; 0043234 : protein complex; 0005634 : nucleus	0006350 : transcription; 0050789 : regulation of biological process; 0006323 : DNA packaging; 0016043 : cell organization and biogenesis; 0006464 : protein modification; 0009987 : cellular process; 0016070 : RNA metabolism; 0007275 : development
Serine/threonine-protein kinase Chk2 (EC 2.7.11.1) (Cds1)	6005850	0016740 : transferase activity; 0043167 : ion binding; 0000166 : nucleotide binding; 0005515 : protein binding	0005634 : nucleus	0007049 : cell cycle; 0050789 : regulation of biological process; 0050896 : response to stimulus; 0006464 : protein modification; 0006793 : phosphorus metabolism
HLA class I histocompatibility antigen, A-68 alpha chain precursor (MHC class I antigen A*68) (Aw-68) (A-28)	6138770		0016020 : membrane; 0043234 : protein complex	0008150 : biological_process; 0050896 : response to stimulus

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Conserved oligomeric Golgi complex component 2 (Low density lipoprotein receptor defect C-complementing protein)	6678676	0005515 : protein binding; 0005215 : transporter activity	0016020 : membrane; 0005794 : Golgi apparatus; 0043234 : protein complex	0015031 : protein transport; 0046907 : intracellular transport; 0006464 : protein modification; 0016192 : vesicle-mediated transport; 0045045 : secretory pathway; 0005975 : carbohydrate metabolism; 0006810 : transport; 0006996 : organelle organization and biogenesis
Component of oligomeric golgi complex 2	6678676		0016020 : membrane	0015031 : protein transport; 0006996 : organelle organization and biogenesis
Golgin subfamily A member 4 (Trans-Golgi p230) (256 kDa golgin) (Golgin-245) (Protein 72.1)	6715600		0000267 : cell fraction; 0005794 : Golgi apparatus	0016192 : vesicle-mediated transport
GOLGA4 protein	6715600			
CASP8 associated protein 2	6912288			
Leucine-rich repeat transmembrane protein FLRT2 precursor (Fibronectin-like domain-containing leucine-rich transmembrane protein 2)	7019381	0005515 : protein binding; 0004871 : signal transducer activity	0016020 : membrane; 0005576 : extracellular region; 0031012 : extracellular matrix	0007155 : cell adhesion
Bleomycin hydrolase (EC 3.4.22.40) (BLM hydrolase) (BMH) (BH)	7245509	0016787 : hydrolase activity	0005737 : cytoplasm; 0000267 : cell fraction; 0005634 : nucleus	0050896 : response to stimulus; 0019538 : protein metabolism
Brain-specific angiogenesis inhibitor 3 precursor	4502357	0004871 : signal transducer activity	0016020 : membrane; 0009986 : cell surface	0007165 : signal transduction
Centaurin-beta 1 (Cnt-b1)	7661880	0030234 : enzyme regulator activity; 0016787 : hydrolase activity; 0043167 : ion binding		0007165 : signal transduction; 0008150 : biological_process

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Centaurin, beta 1	7661880	0030234 : enzyme regulator activity; 0016787 : hydrolase activity; 0043167 : ion binding		0007165 : signal transduction; 0008150 : biological_process
Rho GTPase-activating protein 25	7661882	0030234 : enzyme regulator activity	0005622 : intracellular	0007165 : signal transduction
Protein FAM38A	7662014		0016020 : membrane	
ARHGAP26 protein	7662208		0005622 : intracellular	0007165 : signal transduction
LIM domain and actin-binding protein 1 (Epithelial protein lost in neoplasm)	7705373	0005515 : protein binding; 0043167 : ion binding	0016020 : membrane; 0005856 : cytoskeleton	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0000902 : cellular morphogenesis
UPB1 protein	7706509			
Beta-ureidopropionase (EC 3.5.1.6) (Beta-alanine synthase) (N- carbamoyl-beta-alanine amidohydrolase) (BUP-1)	7706509	0003824 : catalytic activity; 0016787 : hydrolase activity; 0043167 : ion binding		0006807 : nitrogen compound metabolism
BK445C9.6 protein	8393259	0003676 : nucleic acid binding	0005622 : intracellular	
Tuftelin-interacting protein 11	8393259	0003676 : nucleic acid binding; 0005515 : protein binding	0005634 : nucleus; 0005622 : intracellular; 0005681 : spliceosome complex	0030198 : extracellular matrix organization and biogenesis; 0007275 : development; 0006396 : RNA processing; 0009987 : cellular process; 0045045 : secretory pathway

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Mitogen-activated protein kinase kinase kinase 1 (EC 2.7.11.25) (MAPK/ERK kinase kinase 1) (MEK kinase 1) (MEKK 1)	8488988	0016740 : transferase activity; 0043167 : ion binding; 0000166 : nucleotide binding; 0004871 : signal transducer activity; 0005515 : protein binding		0006464 : protein modification; 0006793 : phosphorus metabolism; 0006996 : organelle organization and biogenesis; 0007165 : signal transduction; 0008219 : cell death; 0006928 : cell motility; 0050789 : regulation of biological process; 0016043 : cell organization and biogenesis; 0019538 : protein metabolism; 0050896 : response to stimulus
Tyrosine protein-kinase STYK1 (EC 2.7.10.2) (Serine/threonine/tyrosine kinase 1) (Novel oncogene with kinase domain) (Protein PK-unique)	8922179	0016740 : transferase activity; 0000166 : nucleotide binding	0016020 : membrane	0006464 : protein modification; 0006793 : phosphorus metabolism
F-box only protein 6 (F-box/G-domain protein 2)	8922188	0016874 : ligase activity; 0005515 : protein binding; 0030246 : carbohydrate binding		0019538 : protein metabolism; 0006464 : protein modification
Triple functional domain protein (EC 2.7.11.1) (PTPRF-interacting protein)	8928460	0030234 : enzyme regulator activity; 0016740 : transferase activity; 0000166 : nucleotide binding	0005622 : intracellular	0006464 : protein modification; 0006793 : phosphorus metabolism; 0007165 : signal transduction; 0050789 : regulation of biological process
Alpha-1-acid glycoprotein 1 precursor (AGP 1) (Orosomucoid-1) (OMD 1)	9257232	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus
Orosomucoid 1	9257232	0005488 : binding		

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
ATP-binding cassette sub-family B member 10, mitochondrial precursor (ATP-binding cassette transporter 10) (ABC transporter 10 protein) (Mitochondrial ATP-binding cassette 2) (M-ABC2)	9961244	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane; 0005739 : mitochondrion; 0043190 : ATP-binding cassette (ABC) transporter complex	0006810 : transport
U4/U6.U5 tri-snRNP-associated protein 1 (U4/U6.U5 tri-snRNP-associated 110 kDa protein) (Squamous cell carcinoma antigen recognized by T cells 1) (SART-1) (hSART-1) (hSnu66)	10863889		0005737 : cytoplasm; 0005681 : spliceosome complex; 0005634 : nucleus	0007275 : development; 0009987 : cellular process; 0050789 : regulation of biological process; 0006396 : RNA processing; 0008219 : cell death; 0007049 : cell cycle
Neurogenic differentiation factor 4 (NeuroD4) (Protein atonal homolog 3)	10863999	0003676 : nucleic acid binding; 0030528 : transcription regulator activity	0005634 : nucleus	0007275 : development; 0009987 : cellular process; 0006928 : cell motility; 0007165 : signal transduction; 0008283 : cell proliferation; 0006350 : transcription; 0050789 : regulation of biological process; 0016070 : RNA metabolism
Insulin-like growth factor IA precursor (IGF-IA) (Somatomedin C) (Mechano growth factor) (MGF)	11024682	0005515 : protein binding	0005576 : extracellular region	0006082 : organic acid metabolism; 0007165 : signal transduction; 0007275 : development; 0006260 : DNA replication; 0006928 : cell motility; 0008150 : biological_process; 0008283 : cell proliferation; 0050789 : regulation of biological process
Insulin-like growth factor I (Somatomedin C)	11024682	0005515 : protein binding	0005576 : extracellular region	

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Vesicle-fusing ATPase (EC 3.6.4.6) (Vesicular-fusion protein NSF) (N- ethylmaleimide sensitive fusion protein) (NEM-sensitive fusion protein)	11079228	0043167 : ion binding; 0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005515 : protein binding	0005783 : endoplasmic reticulum	0015031 : protein transport; 0006810 : transport
Calcineurin B homologous protein 2 (Hepatocellular carcinoma- associated antigen 520)	11545811	0043167 : ion binding		
Fibrinogen alpha chain precursor [Contains: Fibrinopeptide A]	11761629	0005515 : protein binding	0000267 : cell fraction; 0005576 : extracellular region; 0043234 : protein complex	0050817 : coagulation; 0050896 : response to stimulus; 0007165 : signal transduction; 0009987 : cellular process; 0019538 : protein metabolism; 0008150 : biological_process; 0008283 : cell proliferation; 0050789 : regulation of biological process
Fibrinogen alpha chain (Fibrinogen, alpha chain, isoform alpha preproprotein)	11761629	0005515 : protein binding	0005576 : extracellular region; 0043234 : protein complex	0007165 : signal transduction; 0009987 : cellular process; 0050817 : coagulation; 0050896 : response to stimulus; 0019538 : protein metabolism
Islet cell autoantigen 1 (69 kDa islet cell autoantigen) (ICA69) (p69) (Islet cell autoantigen p69) (ICAp69)	12545395		0016020 : membrane; 0005737 : cytoplasm; 0016023 : cytoplasmic membrane-bound vesicle; 0005794 : Golgi apparatus	0007267 : cell-cell signaling; 0045045 : secretory pathway; 0050789 : regulation of biological process; 0006810 : transport; 0006836 : neurotransmitter transport
Afadin (Protein AF-6)	12644018	0003774 : motor activity; 0005515 : protein binding	0005856 : cytoskeleton; 0043234 : protein complex; 0016020 : membrane	0007165 : signal transduction; 0007155 : cell adhesion; 0007267 : cell-cell signaling

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Glucosamine-6-phosphate isomerase (EC 3.5.99.6) (Glucosamine-6- phosphate deaminase) (GNPDA) (GlcN6P deaminase) (Oscillin)	13027378	0016787 : hydrolase activity; 0016853 : isomerase activity		0006091 : generation of precursor metabolites and energy; 0005975 : carbohydrate metabolism; 0009308 : amine metabolism; 0008150 : biological_process
ATP-binding cassette sub-family B member 8, mitochondrial precursor (Mitochondrial ATP-binding cassette 1) (M-ABC1)	13123950	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane; 0043190 : ATP-binding cassette (ABC) transporter complex; 0005739 : mitochondrion; 0000267 : cell fraction	0006810 : transport
Myosin-11 (Myosin heavy chain 11) (Myosin heavy chain, smooth muscle isoform) (SMMHC)	13124879	0003774 : motor activity; 0005198 : structural molecule activity; 0000166 : nucleotide binding; 0005515 : protein binding	0005737 : cytoplasm; 0005856 : cytoskeleton; 0043234 : protein complex	0007275 : development; 0009987 : cellular process; 0006461 : protein complex assembly; 0008150 : biological_process; 0030198 : extracellular matrix organization and biogenesis
Myosin, heavy chain 11, smooth muscle	13124879	0003774 : motor activity; 0000166 : nucleotide binding	0005856 : cytoskeleton; 0043234 : protein complex	
Smooth muscle myosin heavy chain isoform SM2	13124879	0003774 : motor activity; 0000166 : nucleotide binding	0005856 : cytoskeleton; 0043234 : protein complex	
Collagen alpha-2(XI) chain precursor	13432104	0005198 : structural molecule activity	0005737 : cytoplasm; 0005576 : extracellular region; 0031012 : extracellular matrix	0030198 : extracellular matrix organization and biogenesis; 0007155 : cell adhesion; 0006811 : ion transport; 0007275 : development; 0008150 : biological_process

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Protein Wnt-5b precursor	14249180	0004871 : signal transducer activity; 0005515 : protein binding	0005576 : extracellular region	0007165 : signal transduction; 0007275 : development
Ig gamma-1 chain C region	14277820	0005515 : protein binding; 0003823 : antigen binding	0000267 : cell fraction	0050896 : response to stimulus
Plasma membrane calcium-transporting ATPase 4 (EC 3.6.3.8) (PMCA4) (Plasma membrane calcium pump isoform 4) (Plasma membrane calcium ATPase isoform 4) (Matrix-remodelling-associated protein 1)	14286105	0003824 : catalytic activity; 0043167 : ion binding; 0000166 : nucleotide binding; 0005515 : protein binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane	0008152 : metabolism; 0006810 : transport; 0006811 : ion transport
General transcription factor II-I (GTFII-I) (TFII-I) (Bruton tyrosine kinase-associated protein 135) (BTK-associated protein 135) (BAP-135) (SRF-Phox1-interacting protein) (SPIN) (Williams-Beuren syndrome chromosome region 6 protein)	14670356	0003676 : nucleic acid binding; 0030528 : transcription regulator activity; 0005515 : protein binding	0005634 : nucleus	0007165 : signal transduction; 0006350 : transcription; 0016043 : cell organization and biogenesis; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Cullin-5 (CUL-5) (Vasopressin-activated calcium-mobilizing receptor) (VACM-1)	14917099	0004871 : signal transducer activity; 0005515 : protein binding; 0005215 : transporter activity		0008283 : cell proliferation; 0050789 : regulation of biological process; 0007049 : cell cycle; 0006464 : protein modification; 0008219 : cell death
Zinc finger imprinted 3	16418391	0003676 : nucleic acid binding; 0043167 : ion binding	0005622 : intracellular; 0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Krueppel-like factor 8 (Basic krueppel-like factor 3) (Zinc finger protein 741)	17366251	0003676 : nucleic acid binding; 0043167 : ion binding	0005622 : intracellular; 0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3) (Beta-IV spectrin)	17368942	0005515 : protein binding; 0005198 : structural molecule activity	0016020 : membrane; 0005737 : cytoplasm; 0005634 : nucleus; 0043005 : neuron projection; 0005856 : cytoskeleton; 0043025 : cell soma	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0007267 : cell-cell signaling; 0016192 : vesicle-mediated transport; 0008150 : biological_process
Monocarboxylate transporter 3 (MCT 3) (Solute carrier family 16 member 8)	17433295	0005215 : transporter activity	0016020 : membrane; 0000267 : cell fraction	0006810 : transport; 0015849 : organic acid transport; 0006811 : ion transport
Enhancer of zeste homolog 1 (Drosophila)	19923202	0003676 : nucleic acid binding	0005634 : nucleus	
Enhancer of zeste homolog 1 (ENX-2)	19923202	0003676 : nucleic acid binding; 0005488 : binding	0005634 : nucleus	0008150 : biological_process; 0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Oxysterol-binding protein (Fragment)	20070331			0006869 : lipid transport; 0006629 : lipid metabolism; 0006810 : transport
Oxysterol-binding protein-related protein 9 (OSBP-related protein 9) (ORP-9)	20070331			0006869 : lipid transport; 0006629 : lipid metabolism; 0006810 : transport
Desmuslin	20137613	0005198 : structural molecule activity; 0005515 : protein binding	0005856 : cytoskeleton	
Prestin (Solute carrier family 26 member 5)	20139418	0003774 : motor activity; 0005215 : transporter activity	0016020 : membrane	0000902 : cellular morphogenesis; 0050789 : regulation of biological process; 0006810 : transport; 0006811 : ion transport; 0008150 : biological_process

Catalogue of human proteins using GO terminology cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Kinesin-like protein KIF2C (Mitotic centromere-associated kinesin) (MCAK) (Kinesin-like protein 6)	20141607	0003774 : motor activity; 0000166 : nucleotide binding; 0005515 : protein binding; 0003676 : nucleic acid binding	0005856 : cytoskeleton; 0043234 : protein complex; 0005634 : nucleus; 0005694 : chromosome	0007049 : cell cycle; 0016043 : cell organization and biogenesis; 0008283 : cell proliferation; 0046907 : intracellular transport
Ribosomal protein S6 kinase alpha-1 (EC 2.7.11.1) (S6K-alpha 1) (90 kDa ribosomal protein S6 kinase 1) (p90-RSK 1) (Ribosomal S6 kinase 1) (RSK-1) (pp90RSK1) (p90S6K) (MAP kinase-activated protein kinase 1a) (MAPKAPK1A)	20149547	0016740 : transferase activity; 0043167 : ion binding; 0000166 : nucleotide binding; 0005515 : protein binding		0007165 : signal transduction; 0006464 : protein modification; 0006793 : phosphorus metabolism

APPENDIX VIII List of *falciparum* proteins identified from plasma with Gene Ontology (GO). The GO Slim ontologies were created using PIR

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Transmission-blocking target antigen S230 precursor	PFB0405w		0016020 : membrane	
Transmission-blocking target antigen S230 precursor	PFB0405w		0016020 : membrane	
Hypothetical protein PFB0555c	PFB0555c			
Erythrocyte membrane protein 1 (PfEMP1)	PFB1055c	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein PFB0770c	PFB0770c			
Hypothetical protein PFB0760w	PFB0760w			
Hypothetical protein PFB0925w	PFB0925w	0005515 : protein binding		0006457 : protein folding
Hypothetical protein PFB0205c	PFB0205c			
Putative transmembrane amino acid transporter protein (Transmembrane amino acid transporter protein, putative)	PFF1430c		0016020 : membrane	
DNA polymerase epsilon, catalytic subunit a, putative (Putative dna polymerase epsilon, catalytic subunit a) (EC 2.7.7.7)	PFF1470c	0016740 : transferase activity; 0003676 : nucleic acid binding; 0016787 : hydrolase activity; 0000166 : nucleotide binding		0006260 : DNA replication
Hypothetical protein PFB0775w	PFB0775w			
Cysteine protease, putative	PFB0350c	0016787 : hydrolase activity		0019538 : protein metabolism
Polyprenyl synthetase, putative	PFB0130w			0006629 : lipid metabolism

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Plasmodium falciparum chromosome 6, complete sequence; segment 4/5	PFF1365c	0016874 : ligase activity	0005622 : intracellular	0006464 : protein modification
Ser/Thr protein kinase, putative	PFB0665w	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Protein kinase, putative	PFB0150c	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Putative SET-domain protein (Set-domain protein, putative)	PFF1440w	0005515 : protein binding; 0043167 : ion binding	0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Erythrocyte membrane protein 3	PFB0095c			
Erythrocyte membrane protein 1 (PfEMP1)	PFB0010w	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Putative sec14-like cytosolic factor or phosphatidylinositol/phosphatidylcholine transfer protein (Sec14-like cytosolic factor or phosphatidylinositol/phosphatidylcholine transfer protein, putative)	PFF1450w			
Hypothetical protein	PFF1490w	0003824 : catalytic activity		0006082 : organic acid metabolism; 0006725 : aromatic compound metabolism; 0046483 : heterocycle metabolism; 0051186 : cofactor metabolism

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Putative transcription or splicing factor-like protein (Transcription or splicing factor-like protein, putative)	PFF1135w	0003676 : nucleic acid binding; 0043167 : ion binding		
Iswi protein homologue	PFF1185w	0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0005515 : protein binding; 0003824 : catalytic activity; 0043167 : ion binding		0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Holo-(Acyl-carrier protein) synthase, putative (EC 2.7.8.7)	PFD0980w	0016740 : transferase activity; 0043167 : ion binding		0006082 : organic acid metabolism; 0006629 : lipid metabolism; 0043170 : macromolecule metabolism
Hypothetical protein	PFF0130c			
Hypothetical protein	PFF0195c			
Hypothetical protein	PFF0470w			
Putative ATP-dependent DEAD box helicase (Atp-dependent dead box helicase, putative)	PFF1140c	0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity		
Hypothetical protein	PFF0670w			
Hypothetical protein	PFF0720w			
Hypothetical protein	PFF1120c			
Hypothetical protein	PFF0145w			
Hypothetical protein	PFF0990c			
Hypothetical protein	PFF1355w	0016787 : hydrolase activity	0030529 : ribonucleoprotein complex	0006396 : RNA processing

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Erythrocyte membrane protein 1 (PfEMP1)	PFD1235w	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Putative oxidoreductase, short-chain dehydrogenase family (Oxidoreductase, short-chain dehydrogenase family, putative) (EC 1.-.- .-)	PFF1265w	0016491 : oxidoreductase activity		0008152 : metabolism
Putative Cdc2-like protein kinase (Cdc2-like protein kinase, putative) (EC 2.7.1.-)	PFF0750w	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
CAMP-dependent protein kinase regulatory subunit, putative	PFL1110c	0030234 : enzyme regulator activity; 0016740 : transferase activity	0005622 : intracellular; 0043234 : protein complex	0006464 : protein modification; 0006793 : phosphorus metabolism
Hypothetical protein (Hypothetical protein, conserved)	PFF0485c	0043167 : ion binding		
DNA polymerase delta catalytic subunit	PF10_0165	0016740 : transferase activity; 0003676 : nucleic acid binding; 0016787 : hydrolase activity; 0000166 : nucleotide binding	0005634 : nucleus	0006260 : DNA replication
Hypothetical protein	PFF1055c			
Hypothetical protein	PFF1280w			
Dolichyl-diphosphooligosaccharide--protein-glycot ransferase,putative (EC 2.4.1.119)	PFI0960w	0016740 : transferase activity	0005783 : endoplasmic reticulum; 0016020 : membrane	0006464 : protein modification
Hypothetical protein PFI1150w	PFI1150w			
RIFIN	PFD0050w			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein PFD0160w	PFD0160w	0003676 : nucleic acid binding; 0043167 : ion binding	0005622 : intracellular	
Bifunctional dihydrofolate reductase-thymidylate synthase (EC 1.5.1.3) (EC 2.1.1.45)	PFD0805w	0016491 : oxidoreductase activity; 0016740 : transferase activity		0009117 : nucleotide metabolism; 0006730 : one-carbon compound metabolism; 0006520 : amino acid metabolism
Hypothetical protein PFD0900w	PFD0900w			
Hypothetical protein PFA0665w	PFA0665w	0004871 : signal transducer activity	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein PFD0190w	PFD0190w		0005634 : nucleus	
Hypothetical protein PFD0840w	PFD0840w			
Hypothetical protein PFI1265w	PFI1265w	0003676 : nucleic acid binding		
Hypothetical protein PFI1305w	PFI1305w			
Hypothetical protein PFD0935c	PFD0935c			
Hypothetical protein PFI1385c	PFI1385c			
Hypothetical protein PFI1410c	PFI1410c			
Hypothetical protein PFD0200c	PFD0200c			
Hypothetical protein PFI1500w	PFI1500w			
DNA replication licensing factor, putative	PFD0790c	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding		0006260 : DNA replication
Plasmodium falciparum protein kinase, putative	PFD0740w	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Hypothetical protein PFI0975c	PFI0975c			
Hypothetical protein PFA0735w	PFA0735w			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Chromosome associated protein, putative	PFD0685c	0000166 : nucleotide binding; 0005515 : protein binding; 0016787 : hydrolase activity	0016020 : membrane; 0005694 : chromosome	0006259 : DNA metabolism; 0006323 : DNA packaging
DNA-directed RNA polymerase (EC 2.7.7.6)	PFE0465c	0016740 : transferase activity; 0003676 : nucleic acid binding	0005634 : nucleus	0006350 : transcription
Hypothetical protein PFE0640w	PFE0640w			
Hypothetical protein PFE1070c	PFE1070c	0005515 : protein binding; 0043167 : ion binding		0006464 : protein modification
Ubiquitin-conjugating enzyme, putative	PFE1350c	0016874 : ligase activity		0006464 : protein modification
Hypothetical protein PFE0230w	PFE0230w			
Hypothetical protein PFE0360c	PFE0360c			
Hypothetical protein PFE1255w	PFE1255w	0000166 : nucleotide binding	0005694 : chromosome	0006259 : DNA metabolism
Hypothetical protein PFE0890c	PFE0890c			0007165 : signal transduction
Met-10+ like protein, putative	PFI0700c			
Hypothetical protein PFI0805w	PFI0805w	0005515 : protein binding; 0043167 : ion binding		
Erythrocyte membrane protein 1 (PfEMP1), truncated	PFE1640w	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein PFI0295c	PFI0295c	0016787 : hydrolase activity; 0003676 : nucleic acid binding		
Hypothetical protein PFE1120w	PFE1120w			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein PFI0260c	PFI0260c	0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0016787 : hydrolase activity; 0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process; 0046907 : intracellular transport
Hypothetical protein PFE0240c (Fragment)	PFE0240c			
Chromosome condensation protein, putative	PFE0450w	0000166 : nucleotide binding; 0005515 : protein binding	0016020 : membrane; 0005694 : chromosome	0006259 : DNA metabolism; 0006323 : DNA packaging
Diaphanous homolog, putative	PFE1545c	0005488 : binding; 0005515 : protein binding		0016043 : cell organization and biogenesis
Erythrocyte membrane protein 1 (PfEMP1)	PFI0005w	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Rifin	PFI0055c			
Hypothetical protein PFI0495w	PFI0495w			
Hypothetical protein PFI0665w	PFI0665w			
Probable U3 small nucleolar RNA-associated protein 11 (U3 snoRNA-associated protein 11)	PFL2295w		0030529 : ribonucleoprotein complex; 0005634 : nucleus	0006396 : RNA processing; 0042254 : ribosome biogenesis and assembly
Pre-mRNA splicing factor RNA helicase, putative	PFL1525c	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity		
Hypothetical protein	PFL1675c			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Reticulocyte-binding protein, putative	PFL2520w			
Erythrocyte membrane protein 1 (PfEMP1)	PFL1950w	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein	PFL1560c			
Hypothetical protein	PFL2110c			
Hypothetical protein PFE0085c	PFE0085c	0005488 : binding		
Hypothetical protein	PFL2390c			
Hypothetical protein	PFL1930w			
Hypothetical protein	PFL1410c	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane	0006810 : transport
Hypothetical protein	PFL1645w			
Hypothetical protein	PFL1650w			
Hypothetical protein	PFL2085w			
Hypothetical protein	PFL2350c			
CutA, putative	PFL2375c			
ATP-dependent helicase, putative (EC 3.6.1.3)	PFE0205w	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity		
Phenylalanyl-tRNA synthetase alpha chain, putative	PFL1540c	0016874 : ligase activity; 0000166 : nucleotide binding		0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism
3-hydroxyisobutyryl-coenzyme A hydrolase, putative	PFL1940w	0003824 : catalytic activity; 0016787 : hydrolase activity		0008152 : metabolism

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein	PFL2150c			
Erythrocyte membrane protein 1 (PfEMP1)	PFL2665c	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein	PFL0045c			
Hypothetical protein	PFL0405w			
Hypothetical protein PF08_0111	PF08_0111	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity		
Erythrocyte membrane protein 1 (PfEMP1)	PF08_0141	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein	PFL0130c			
Hypothetical protein	PFL0650c			
Hypothetical protein	PFL1330c			
Seryl-tRNA synthetase, putative	PFL0770w	0016874 : ligase activity; 0000166 : nucleotide binding		0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism
Hypothetical protein	PFL1035w			
Hypothetical protein MAL8P1.154	MAL8P1.154	0016491 : oxidoreductase activity		0006091 : generation of precursor metabolites and energy
Hypothetical protein	PFL1280w			
Hypothetical protein	PFL0410w			
Hypothetical protein	PFL0440c	0005515 : protein binding; 0043167 : ion binding		0006464 : protein modification

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein	PFL0610w			
Hypothetical protein	PFL1395c			
Erythrocyte membrane protein 1 (PfEMP1)	PF08_0140	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Rifin	PFL0015c			
Hypothetical protein	PFL0555c			
Signal recognition particle 19 kd protein, putative	PFL0785c	0003676 : nucleic acid binding	0005737 : cytoplasm; 0030529 : ribonucleoprotein complex	0015031 : protein transport; 0046907 : intracellular transport
Polyadenylate-binding protein, putative	PFL1170w	0000166 : nucleotide binding; 0003676 : nucleic acid binding		
Hypothetical protein	PFL1320w			
Hypothetical protein MAL7P1.138	MAL7P1.138			
Hypothetical protein PF07_0118	PF07_0118			
Hypothetical protein MAL7P1.171	MAL7P1.171			
Hypothetical protein PF08_0080	PF08_0080			
Cg2 protein	PF07_0037			
Erythrocyte membrane protein 1 (PfEMP1)	MAL7P1.50	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein PF07_0067	PF07_0067			
Ubiquitin-protein ligase 1, putative	MAL8P1.23	0016874 : ligase activity	0005737 : cytoplasm; 0005622 : intracellular	0006464 : protein modification; 0015931 : nucleobase\, nucleoside\, nucleotide and nucleic acid transport

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein MAL7P1.120	MAL7P1.120	0043167 : ion binding		
Hypothetical protein MAL7P1.167	MAL7P1.167			
Hypothetical protein PF08_0028	PF08_0028			
Hypothetical protein PF07_0021	PF07_0021	0016787 : hydrolase activity; 0043167 : ion binding		0019538 : protein metabolism; 0009405 : pathogenesis
Erythrocyte membrane protein 1 (PfEMP1)	MAL7P1.187	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein MAL8P1.73	MAL8P1.73			
Hypothetical protein MAL8P1.113	MAL8P1.113	0016787 : hydrolase activity	0005634 : nucleus	0019538 : protein metabolism
Erythrocyte membrane protein 1 (PfEMP1)	PF07_0051	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein PF07_0107	PF07_0107			
Hypothetical protein MAL7P1.146	MAL7P1.146			
Hypothetical protein PF08_0012	PF08_0012	0016740 : transferase activity; 0043167 : ion binding	0005634 : nucleus	0006323 : DNA packaging
Hypothetical protein MAL8P1.29	MAL8P1.29			
Hypothetical protein MAL8P1.124	MAL8P1.124			
Ser/Thr protein kinase (EC 2.7.1.37)	MAL13P1.278	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Hypothetical protein MAL13P1.228	MAL13P1.228			
Hypothetical protein MAL13P1.247	MAL13P1.247			
Hypothetical protein PF13_0239	PF13_0239			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein MAL13P1.246	MAL13P1.246	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane	0006810 : transport
Hypothetical protein MAL13P1.273	MAL13P1.273			
Hypothetical protein MAL7P1.16	MAL7P1.16			
Hypothetical protein Phat128	PF13_0172			
Hypothetical protein MAL13P1.201	MAL13P1.201			
Aspartate carbamoyltransferase (EC 2.1.3.2)	MAL13P1.221	0043176 : amine binding; 0016740 : transferase activity		0006520 : amino acid metabolism; 0009117 : nucleotide metabolism; 0009112 : nucleobase metabolism
Hypothetical protein MAL13P1.352	MAL13P1.352			
Hypothetical protein PF07_0014	PF07_0014			
Hypothetical protein PF13_0167	PF13_0167			
Hypothetical protein PF13_0189	PF13_0189			
Hypothetical protein MAL13P1.202	MAL13P1.202			
Phosphatidylinositol transfer protein, putative	MAL13P1.256		0005622 : intracellular	0006810 : transport
Hypothetical protein MAL13P1.234	MAL13P1.234			
Hypothetical protein PF13_0355	PF13_0355			
Hypothetical protein MAL7P1.15	MAL7P1.15			
Hypothetical protein MAL7P1.17	MAL7P1.17			
Alanine--tRNA ligase, putative (EC 6.1.1.7)	PF13_0354	0016874 : ligase activity; 0000166 : nucleotide binding		0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Erythrocyte membrane protein 1 (PfEMP1)	PFD1005c	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein PF13_0078	PF13_0078			
Helicase, putative	MAL13P1.134	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0003824 : catalytic activity	0005634 : nucleus	0006139 : nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism
Hypothetical protein	PF11_0511			
Erythrocyte membrane protein 1 (PfEMP1)	PFD0995c	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Erythrocyte membrane-associated antigen, putative	PFD1045c			
Hypothetical protein MAL13P1.23	MAL13P1.23	0005215 : transporter activity	0016020 : membrane	0006811 : ion transport
Hypothetical protein PF13_0072	PF13_0072			
Hypothetical protein PF13_0126	PF13_0126		0005622 : intracellular	0044237 : cellular metabolism
Hypothetical protein MAL13P1.147	MAL13P1.147			
Alpha-tubulin ii	PFD1050w	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0005198 : structural molecule activity	0043234 : protein complex; 0005856 : cytoskeleton	0019538 : protein metabolism; 0016043 : cell organization and biogenesis; 0046907 : intracellular transport
Hypothetical protein MAL13P1.13	MAL13P1.13			
Hypothetical protein MAL13P1.133	MAL13P1.133			
Hypothetical protein MAL13P1.32	MAL13P1.32			
Hypothetical protein PF13_0035	PF13_0035			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Peptidyl-prolyl cis-trans isomerase	MAL13P1.68	0016853 : isomerase activity		0006457 : protein folding
Hypothetical protein PF13_0161	PF13_0161			
Hypothetical protein PFD1160w	PFD1160w			
Hypothetical protein MAL13P1.19	MAL13P1.19	0016740 : transferase activity		
DNA ligase 1 (EC 6.5.1.1)	MAL13P1.22	0016874 : ligase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding		0006260 : DNA replication; 0006310 : DNA recombination; 0006281 : DNA repair
Hypothetical protein PF13_0060	PF13_0060			
Hypothetical protein	PF11_0185			
Hypothetical protein	PF11_0153			
Hypothetical protein	PF11_0226			
Dynein heavy chain, putative	PF11_0240	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0046907 : intracellular transport
Hypothetical protein	PF11_0297	0030528 : transcription regulator activity	0005634 : nucleus	0006350 : transcription; 0050789 : regulation of biological process
Hypothetical protein	PF11_0326			
Malaria antigen	PF11_0404			
Hypothetical protein	PF11_0468			
Threonine--tRNA ligase, putative	PF11_0270	0016874 : ligase activity; 0000166 : nucleotide binding		0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism
Hypothetical protein	PF11_0374	0003676 : nucleic acid binding		
Hypothetical protein	PF11_0408			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein	PF11_0158			
Hypothetical protein	PF11_0392			
DNA-dependent RNA polymerase	PF11_0264	0016740 : transferase activity; 0003676 : nucleic acid binding		0006350 : transcription
Hypothetical protein	PF11_0528	0016787 : hydrolase activity; 0043167 : ion binding		0019538 : protein metabolism; 0009405 : pathogenesis
Hypothetical protein	PF11_0347	0003676 : nucleic acid binding		
Hypothetical protein	PF11_0127	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Hypothetical protein	PF11_0129			
Hypothetical protein	PF11_0207			
Hypothetical protein	PF11_0417			
Hypothetical protein	PF11_0481	0016874 : ligase activity		0006464 : protein modification
Hypothetical protein	PF11_0092		0016020 : membrane	
Hypothetical protein	PF10_0179			
Hypothetical protein	PF10_0262			
Hypothetical protein	PF10_0342			
Hypothetical protein	PF11_0077	0000166 : nucleotide binding; 0003676 : nucleic acid binding; 0003824 : catalytic activity		
Asparagine-rich antigen	PF10_0075			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Dynein heavy chain, putative	PF10_0224	0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0046907 : intracellular transport
Hypothetical protein	PF10_0241			
Hypothetical protein	PF10_0251			
Hypothetical protein	PF10_0180			
Hypothetical protein	PF10_0215			
Hypothetical protein	PF10_0267			
Hypothetical protein	PF10_0274			
Hypothetical protein	PF11_0057			
Hypothetical protein	PF10_0026			
Hypothetical protein	PF10_0044			
Rifin	PF10_0403			
Hypothetical protein	PF11_0080			
Hypothetical protein	PF10_0117	0003676 : nucleic acid binding; 0005515 : protein binding; 0043167 : ion binding	0005634 : nucleus	0006281 : DNA repair
Hypothetical protein	PF10_0140			
Hypothetical protein	PF10_0189			
Hypothetical protein	PF14_0435			
Hypothetical protein	PF14_0470	0016787 : hydrolase activity; 0003676 : nucleic acid binding		0006259 : DNA metabolism
Dynein beta chain, putative	PF14_0626	0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0046907 : intracellular transport

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Erythrocyte membrane protein 1 (PfEMP1)	PF10_0001	0004871 : signal transducer activity; 0030246 : carbohydrate binding	0016020 : membrane	0009405 : pathogenesis
Hypothetical protein	PF14_0419	0016787 : hydrolase activity; 0043167 : ion binding		0019538 : protein metabolism; 0009405 : pathogenesis
Hypothetical protein	PF14_0480			0007155 : cell adhesion; 0007275 : development
Hypothetical protein	PF14_0507			
3'-5' exonuclease, putative	PF14_0473	0016787 : hydrolase activity; 0003676 : nucleic acid binding	0005622 : intracellular	
Hypothetical protein	PF14_0538			
Hypothetical protein	PF14_0509			
Hypothetical protein	PF14_0631			
Hypothetical protein	PF14_0668			
Hypothetical protein	PF14_0504			
Ferlin, putative	PF14_0530			
Heat shock protein, putative	PF14_0417	0000166 : nucleotide binding; 0005515 : protein binding		0006457 : protein folding; 0050896 : response to stimulus
Hypothetical protein	PF14_0405			
Hypothetical protein	PF14_0454			
Hypothetical protein	PF14_0478	0016491 : oxidoreductase activity; 0000166 : nucleotide binding; 0048037 : cofactor binding		
Hypothetical protein	PF14_0490			
Hypothetical protein	PF14_0404			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein	PF10_0021			
Hypothetical protein	PF14_0165			
Hypothetical protein	PF14_0379			
Hypothetical protein	PF14_0263			
Hypothetical protein	PF14_0402			
Hypothetical protein	PF14_0175			
Dynein-associated protein, putative	PF14_0202			
NAD(P)H-dependent glutamate synthase, putative	PF14_0334	0003824 : catalytic activity; 0043167 : ion binding; 0016491 : oxidoreductase activity	0016020 : membrane	0006091 : generation of precursor metabolites and energy; 0008152 : metabolism; 0006520 : amino acid metabolism; 0006807 : nitrogen compound metabolism
Hypothetical protein	PF14_0383	0003676 : nucleic acid binding; 0043167 : ion binding	0005622 : intracellular	
Hypothetical protein	PF14_0190			
Ser/Thr protein kinase, putative	PF14_0392	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Hypothetical protein	PF14_0170			
Hypothetical protein	PF14_0345	0000166 : nucleotide binding	0005622 : intracellular	
Aspartyl protease, putative	PF14_0281	0016787 : hydrolase activity		0019538 : protein metabolism
Hypothetical protein	PF14_0155	0016740 : transferase activity		0009058 : biosynthesis

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Coatamer protein, beta subunit, putative	PF14_0277	0005488 : binding	0005794 : Golgi apparatus; 0016020 : membrane; 0016023 : cytoplasmic membrane-bound vesicle	0015031 : protein transport; 0046907 : intracellular transport; 0006810 : transport
Hypothetical protein	PF14_0293			
Cleavage and polyadenylation specifity factor protein, putative	PF14_0364	0016787 : hydrolase activity		
Hypothetical protein	PF14_0372			
Hypothetical protein	PF14_0282	0016787 : hydrolase activity		
Hypothetical protein	PF14_0318			
Hypothetical protein	PF14_0326	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005515 : protein binding	0005622 : intracellular	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Hypothetical protein	PF14_0059			
Hypothetical protein	PF14_0073			
Hypothetical protein	PF14_0074			
Rhoptry-associated protein 1	PF14_0102			
Hypothetical protein PFD0535w	PFD0535w			
Hypothetical protein PFC0221c	PFC0221c			
DNA repair helicase, putative	PF14_0081	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0003824 : catalytic activity		0006139 : nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism
Hypothetical protein	PF14_0082			

Appendix VIII Cont'd

Protein Name	Matched Accesion Number	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein PFD0430c	PFD0430c			
Hypothetical protein PFD0440w	PFD0440w	0016787 : hydrolase activity; 0043167 : ion binding		
Hypothetical protein	PF14_0088	0016491 : oxidoreductase activity		
Hypothetical protein	PF14_0084			
Hypothetical protein	PF14_0101		0005634 : nucleus	0007049 : cell cycle; 0007059 : chromosome segregation; 0006396 : RNA processing; 0051301 : cell division
Hypothetical protein PFD0320c (Fragment)	PFD0320c			
POM1, putative	PF14_0112	0016740 : transferase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0016787 : hydrolase activity; 0003824 : catalytic activity	0005622 : intracellular	0006260 : DNA replication; 0006304 : DNA modification

APPENDIX IX List of CSF human proteins with Gene Ontologies

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Rod cGMP-specific 3',5'-cyclic phosphodiesterase subunit alpha (EC 3.1.4.35) (GMP-PDE alpha) (PDE V-B1)	105117	0016787 : hydrolase activity; 0003824 : catalytic activity	0016020 : membrane	0008150 : biological_process; 0007165 : signal transduction; 0050896 : response to stimulus
Alpha-1-antichymotrypsin precursor (ACT) [Contains: Alpha-1- antichymotrypsin His-Pro-less]	112874	0003676 : nucleic acid binding; 0030234 : enzyme regulator activity; 0005515 : protein binding	0005622 : intracellular; 0005576 : extracellular region; 0005634 : nucleus	0006629 : lipid metabolism; 0050789 : regulation of biological process; 0050896 : response to stimulus
Ig alpha-1 chain C region	113584	0003823 : antigen binding		0006464 : protein modification; 0050896 : response to stimulus
Arylamine N-acetyltransferase 1 (EC 2.3.1.5) (Arylamide acetylase 1) (Monomorphic arylamine N-acetyltransferase) (MNAT) (N-acetyltransferase type 1) (NAT-1)	114234	0016740 : transferase activity		0008152 : metabolism
60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock protein 60) (HSP-60) (Mitochondrial matrix protein P1) (P60 lymphocyte protein) (HuCHA60)	129379	0000166 : nucleotide binding; 0005515 : protein binding	0005737 : cytoplasm; 0005739 : mitochondrion	0006457 : protein folding; 0015031 : protein transport; 0046907 : intracellular transport; 0019538 : protein metabolism; 0050896 : response to stimulus; 0008219 : cell death; 0050789 : regulation of biological process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Prostaglandin G/H synthase 1 precursor (EC 1.14.99.1) (Cyclooxygenase- 1) (COX-1) (Prostaglandin-endoperoxide synthase 1) (Prostaglandin H2 synthase 1) (PGH synthase 1) (PGHS-1) (PHS 1)	129899	0016491 : oxidoreductase activity; 0043167 : ion binding; 0016209 : antioxidant activity	0016020 : membrane; 0005737 : cytoplasm; 0000267 : cell fraction; 0005634 : nucleus	0006629 : lipid metabolism; 0009987 : cellular process; 0006082 : organic acid metabolism; 0008150 : biological_process
Probable global transcription activator SNF2L1 (EC 3.6.1.-) (Nucleosome remodeling factor subunit SNF2L) (ATP-dependent helicase SMARCA1) (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 1)	134584	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0030528 : transcription regulator activity; 0005515 : protein binding; 0003824 : catalytic activity	0005634 : nucleus	0007275 : development; 0006323 : DNA packaging; 0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process; 0009987 : cellular process
Sperm activating protein subunit I, apolipoprotein A1, SPAP subunit I (Fragment)	235865			
Transthyretin precursor (Prealbumin) (TBPA) (TTR) (ATTR)	494652	0005515 : protein binding; 0005496 : steroid binding; 0019840 : isoprenoid binding; 0019842 : vitamin binding; 0005215 : transporter activity	0005576 : extracellular region	0006810 : transport; 0006575 : amino acid derivative metabolism; 0009308 : amine metabolism; 0042445 : hormone metabolism
Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK-1) (Keratin-1) (K1) (67 kDa cytokeratin) (Hair alpha protein)	1346343	0030246 : carbohydrate binding; 0004871 : signal transducer activity; 0005198 : structural molecule activity; 0005515 : protein binding	0016020 : membrane; 0005856 : cytoskeleton	0050789 : regulation of biological process; 0050817 : coagulation; 0050896 : response to stimulus; 0006800 : oxygen and reactive oxygen species metabolism; 0007275 : development; 0019538 : protein metabolism

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Myosin-10 (Myosin heavy chain 10) (Myosin heavy chain, nonmuscle IIb) (Nonmuscle myosin heavy chain IIb) (NMMHC II-b) (NMMHC-IIB) (Cellular myosin heavy chain, type B) (Nonmuscle myosin heavy chain-B) (NMMHC- B)	1346640	0003774 : motor activity; 0000166 : nucleotide binding; 0005515 : protein binding	0005737 : cytoplasm; 0005856 : cytoskeleton; 0043234 : protein complex; 0030496 : midbody; 0005623 : cell	0016044 : membrane organization and biogenesis; 0050896 : response to stimulus; 0000902 : cellular morphogenesis; 0050789 : regulation of biological process; 0006928 : cell motility; 0007275 : development; 0016043 : cell organization and biogenesis; 0016192 : vesicle-mediated transport; 0045045 : secretory pathway
Melanoma inhibitory activity protein 3 precursor (Protein TANGO) (C219-reactive peptide) (D320)	1665825			
Alpha-1-antitrypsin precursor (Alpha-1 protease inhibitor) (Alpha-1- antiproteinase)	1703025	0030234 : enzyme regulator activity; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus
Cathepsin G precursor (EC 3.4.21.20) (CG)	2392230	0016787 : hydrolase activity	0000267 : cell fraction	0019538 : protein metabolism; 0050896 : response to stimulus
Hemoglobin subunit beta (Hemoglobin beta chain) (Beta-globin)	3660145	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005515 : protein binding; 0008430 : selenium binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport; 0006807 : nitrogen compound metabolism; 0050789 : regulation of biological process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Serum albumin precursor	4389275	0003823 : antigen binding; 0005215 : transporter activity; 0003676 : nucleic acid binding; 0019825 : oxygen binding; 0016209 : antioxidant activity; 0008144 : drug binding; 0005515 : protein binding; 0043167 : ion binding; 0019842 : vitamin binding; 0048037 : cofactor binding; 0008289 : lipid binding	0005576 : extracellular region; 0043234 : protein complex	0008219 : cell death; 0050789 : regulation of biological process; 0006810 : transport; 0007154 : cell communication; 0050896 : response to stimulus; 0044419 : interaction between organisms; 0016043 : cell organization and biogenesis; 0051179 : localization
Membrane copper amine oxidase (EC 1.4.3.6) (Semicarbazide-sensitive amine oxidase) (SSAO) (Vascular adhesion protein 1) (VAP-1) (HPAO)	4502119	0016491 : oxidoreductase activity; 0005515 : protein binding; 0043167 : ion binding; 0048037 : cofactor binding	0016020 : membrane; 0009986 : cell surface	0006091 : generation of precursor metabolites and energy; 0007155 : cell adhesion; 0050896 : response to stimulus; 0009308 : amine metabolism
Gelsolin precursor (Actin-depolymerizing factor) (ADF) (Brevin) (AGEL)	4504165	0005515 : protein binding; 0043167 : ion binding	0031252 : leading edge; 0042995 : cell projection; 0005576 : extracellular region; 0005856 : cytoskeleton; 0005737 : cytoplasm	0016043 : cell organization and biogenesis; 0019538 : protein metabolism; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0016192 : vesicle-mediated transport
Gelsolin (Amyloidosis, Finnish type)	4504165	0005515 : protein binding; 0043167 : ion binding	0005737 : cytoplasm; 0005856 : cytoskeleton; 0005576 : extracellular region	0016043 : cell organization and biogenesis; 0019538 : protein metabolism; 0050789 : regulation of biological process; 0051261 : protein depolymerization

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hemoglobin subunit delta (Hemoglobin delta chain) (Delta-globin)	4504351	0005488 : binding; 0043167 : ion binding; 0019825 : oxygen binding; 0005215 : transporter activity; 0046906 : tetrapyrrole binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport
Delta-globin chain (Hemoglobin delta)	4504351	0005488 : binding; 0019825 : oxygen binding	0005737 : cytoplasm; 0043234 : protein complex	0006810 : transport
Alpha-1-acid glycoprotein 2 precursor (AGP 2) (Orosomucoid-2) (OMD 2)	4505529	0005488 : binding	0005576 : extracellular region	0050896 : response to stimulus
Orosomucoid 2	4505529	0005488 : binding		
Protein S100-A9 (S100 calcium-binding protein A9) (Calgranulin-B) (Migration inhibitory factor-related protein 14) (MRP-14) (P14) (Leukocyte L1 complex heavy chain) (Calprotectin L1H subunit)	4506773	0004871 : signal transducer activity; 0043167 : ion binding	0005576 : extracellular region	0007267 : cell-cell signaling; 0050896 : response to stimulus
Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1) (High affinity sodium-glucose cotransporter)	4507031	0043167 : ion binding; 0005215 : transporter activity	0016020 : membrane; 0045177 : apical part of cell; 0042995 : cell projection	0008150 : biological_process; 0008643 : carbohydrate transport; 0006811 : ion transport; 0006810 : transport; 0007275 : development
Sulfotransferase 1C1 (EC 2.8.2.-) (SULT1C#1) (ST1C2) (humSULTC2)	4507305	0016740 : transferase activity	0005737 : cytoplasm	0009308 : amine metabolism
Thyrotroph embryonic factor	4507431	0030528 : transcription regulator activity; 0003676 : nucleic acid binding; 0005515 : protein binding	0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process; 0048511 : rhythmic process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Serine--pyruvate aminotransferase (EC 2.6.1.51) (SPT) (Alanine--glyoxylate aminotransferase) (EC 2.6.1.44) (AGT)	4557289	0016740 : transferase activity; 0005515 : protein binding; 0019842 : vitamin binding; 0048037 : cofactor binding	0005739 : mitochondrion; 0042579 : microbody	0008152 : metabolism; 0006081 : aldehyde metabolism; 0006082 : organic acid metabolism; 0006996 : organelle organization and biogenesis; 0015031 : protein transport; 0046907 : intracellular transport
Hypothetical protein AGXT (AGXT protein)	4557289	0016740 : transferase activity		0008152 : metabolism
Cholinesterase precursor (EC 3.1.1.8) (Acylcholine acylhydrolase) (Choline esterase II) (Butyrylcholine esterase) (Pseudocholinesterase)	4557351	0003824 : catalytic activity; 0016787 : hydrolase activity; 0005515 : protein binding	0005576 : extracellular region; 0005783 : endoplasmic reticulum; 0005634 : nucleus; 0016020 : membrane	0017144 : drug metabolism
Tyrosine-protein kinase BTK (EC 2.7.10.2) (Bruton tyrosine kinase) (Agammaglobulinaemia tyrosine kinase) (ATK) (B cell progenitor kinase) (BPK)	4557377	0016740 : transferase activity; 0000166 : nucleotide binding; 0005515 : protein binding; 0043167 : ion binding	0005737 : cytoplasm; 0043229 : intracellular organelle; 0016020 : membrane	0006464 : protein modification; 0006793 : phosphorus metabolism; 0007275 : development; 0008219 : cell death; 0050789 : regulation of biological process; 0007165 : signal transduction
Jagged-1 precursor (Jagged1) (hJ1) (CD339 antigen)	4557679	0005515 : protein binding; 0043167 : ion binding; 0005198 : structural molecule activity	0016020 : membrane; 0005576 : extracellular region	0007165 : signal transduction; 0050789 : regulation of biological process; 0007275 : development; 0009987 : cellular process; 0007154 : cell communication
Kell blood group glycoprotein (EC 3.4.24.-) (CD238 antigen)	4557691	0016787 : hydrolase activity; 0005515 : protein binding; 0043167 : ion binding	0016020 : membrane; 0005856 : cytoskeleton	0006464 : protein modification; 0019538 : protein metabolism; 0008150 : biological_process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Neurofibromin (Neurofibromatosis-related protein NF-1) [Contains: Neurofibromin truncated]	4557793	0030234 : enzyme regulator activity; 0005515 : protein binding	0005737 : cytoplasm; 0005622 : intracellular	0008283 : cell proliferation; 0050789 : regulation of biological process; 0007165 : signal transduction; 0007049 : cell cycle; 0007275 : development; 0009987 : cellular process
Succinyl-CoA:3-ketoacid-coenzyme A transferase 1, mitochondrial precursor (EC 2.8.3.5) (Somatic-type succinyl CoA:3-oxoacid CoA- transferase) (Scot-S)	4557817	0016740 : transferase activity	0005739 : mitochondrion	0008152 : metabolism; 0005975 : carbohydrate metabolism; 0051186 : cofactor metabolism
Transferrin	4557871	0043167 : ion binding	0005576 : extracellular region; 0016020 : membrane; 0045177 : apical part of cell; 0016023 : cytoplasmic membrane-bound vesicle	0019725 : cell homeostasis; 0006811 : ion transport
Lactotransferrin precursor (EC 3.4.21.-) (Lactoferrin) (Talalactoferrin alfa) [Contains: Kaliocin-1; Lactoferroxin A; Lactoferroxin B; Lactoferroxin C]	4699853	0016787 : hydrolase activity; 0043167 : ion binding	0005576 : extracellular region	0006810 : transport; 0019725 : cell homeostasis; 0050896 : response to stimulus; 0006811 : ion transport
Probable ATP-dependent RNA helicase DDX5 (EC 3.6.1.-) (DEAD box protein 5) (RNA helicase p68)	4758138	0016787 : hydrolase activity; 0000166 : nucleotide binding; 0003824 : catalytic activity; 0003676 : nucleic acid binding	0005681 : spliceosome complex; 0005634 : nucleus	0016049 : cell growth; 0006396 : RNA processing

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
FK506-binding protein 5 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase) (51 kDa FK506-binding protein) (FKBP- 51) (54 kDa progesterone receptor-associated immunophilin) (FKBP54) (P54) (FF1 antigen) (HSP90-binding immunophilin) (Androgen-regulated protein 6)	4758384	0005488 : binding; 0008144 : drug binding; 0016853 : isomerase activity; 0005515 : protein binding	0005634 : nucleus	0006457 : protein folding
FK506 binding protein 5	4758384	0005488 : binding; 0016853 : isomerase activity		0006457 : protein folding
NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (NADH-ubiquinone oxidoreductase 13 kDa-A subunit) (Complex I-13kD-A) (CI-13kD-A)	4758792	0016491 : oxidoreductase activity	0005739 : mitochondrion	0006091 : generation of precursor metabolites and energy; 0006793 : phosphorus metabolism
NDUFS6 protein	4758792			
N-myc-interactor (Nmi) (N-myc and STAT interactor)	4758814	0005515 : protein binding; 0030528 : transcription regulator activity	0005737 : cytoplasm	0050896 : response to stimulus; 0007165 : signal transduction; 0006350 : transcription; 0016070 : RNA metabolism
Sodium/hydrogen exchanger 5 (Na(+)/H(+) exchanger 5) (NHE-5) (Solute carrier family 9 member 5)	4759144	0043167 : ion binding; 0005215 : transporter activity	0016020 : membrane	0006811 : ion transport; 0006810 : transport; 0042592 : homeostasis
Hypothetical protein DKFZp434N222 (Fragment)	4759144			

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
FKBP12-rapamycin complex-associated protein (FK506-binding protein 12- rapamycin complex-associated protein 1) (Rapamycin target protein) (RAPT1) (Mammalian target of rapamycin) (mTOR)	4826730	0016740 : transferase activity; 0005488 : binding; 0005515 : protein binding	0016020 : membrane; 0005737 : cytoplasm; 0043234 : protein complex	0007165 : signal transduction; 0016049 : cell growth; 0007049 : cell cycle; 0050789 : regulation of biological process; 0019538 : protein metabolism; 0006793 : phosphorus metabolism; 0007154 : cell communication; 0050896 : response to stimulus; 0008150 : biological_process
Haptoglobin precursor [Contains: Haptoglobin alpha chain; Haptoglobin beta chain]	4826762	0005515 : protein binding; 0016787 : hydrolase activity	0005576 : extracellular region	0050896 : response to stimulus; 0019725 : cell homeostasis; 0019538 : protein metabolism
Corticosteroid 11-beta-dehydrogenase isozyme 1 (EC 1.1.1.146) (11-DH) (11-beta-hydroxysteroid dehydrogenase 1) (11-beta-HSD1)	5031765	0016491 : oxidoreductase activity	0016020 : membrane; 0000267 : cell fraction; 0005783 : endoplasmic reticulum	0008152 : metabolism; 0006629 : lipid metabolism; 0007275 : development; 0042445 : hormone metabolism
Lipase, endothelial	5174497	0016787 : hydrolase activity; 0003824 : catalytic activity		0006629 : lipid metabolism
Endothelial lipase precursor (EC 3.1.1.3) (Endothelial cell-derived lipase) (EDL) (EL)	5174497	0030246 : carbohydrate binding; 0016787 : hydrolase activity; 0003824 : catalytic activity; 0005515 : protein binding		0006629 : lipid metabolism

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1 (EC 2.8.2.8) (Glucosaminyl N-deacetylase/N-sulfotransferase 1) (NDST- 1) ([Heparan sulfate]-glucosamine N-sulfotransferase 1) (HSNST 1) (N- heparan sulfate sulfotransferase 1) (N-HSST 1) [Includes: Heparan sulfate N-deacetylase 1 (EC 3.-.-.-); Heparan sulfate N-sulfotransferase 1 (EC 2.8.2.-)]	6137498	0003824 : catalytic activity; 0016740 : transferase activity; 0016787 : hydrolase activity	0016020 : membrane	0006464 : protein modification; 0050896 : response to stimulus; 0006790 : sulfur metabolism; 0005975 : carbohydrate metabolism; 0019538 : protein metabolism
CASP8 associated protein 2	6912288			
CASP8-associated protein 2 (FLICE-associated huge protein)	6912288	0005515 : protein binding; 0030234 : enzyme regulator activity; 0030528 : transcription regulator activity	0005737 : cytoplasm; 0005634 : nucleus	0007165 : signal transduction; 0008219 : cell death; 0050789 : regulation of biological process; 0006350 : transcription; 0016070 : RNA metabolism
Zinc finger protein 281 (Zinc finger DNA-binding protein 99) (Transcription factor ZBP-99) (GC-box-binding zinc finger protein 1)	6912752	0003676 : nucleic acid binding; 0030528 : transcription regulator activity; 0043167 : ion binding	0016591 : DNA-directed RNA polymerase II\, holoenzyme; 0030880 : RNA polymerase complex; 0005622 : intracellular; 0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Putative ribosomal RNA methyltransferase 2 (EC 2.1.1.-) (rRNA (uridine-2'-O-)-methyltransferase)	7019377	0016740 : transferase activity	0005634 : nucleus	0006396 : RNA processing; 0042254 : ribosome biogenesis and assembly

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Pescadillo homolog 1	7657455		0005622 : intracellular; 0005730 : nucleolus; 0005634 : nucleus	0008150 : biological_process; 0008283 : cell proliferation
Probable leucyl-tRNA synthetase, mitochondrial precursor (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS)	7661872	0016874 : ligase activity; 0000166 : nucleotide binding	0005739 : mitochondrion	0006412 : protein biosynthesis; 0006520 : amino acid metabolism; 0016070 : RNA metabolism
Lipin-2	7662022		0005634 : nucleus	
Vinculin (Metavinculin)	7669550	0005515 : protein binding; 0016491 : oxidoreductase activity; 0005198 : structural molecule activity	0016020 : membrane; 0005737 : cytoplasm; 0043234 : protein complex; 0005856 : cytoskeleton	0007155 : cell adhesion; 0000902 : cellular morphogenesis; 0006461 : protein complex assembly; 0007028 : cytoplasm organization and biogenesis; 0006928 : cell motility; 0050789 : regulation of biological process
LIM domain and actin-binding protein 1 (Epithelial protein lost in neoplasm)	7705373	0005515 : protein binding; 0043167 : ion binding	0016020 : membrane; 0005856 : cytoskeleton	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0000902 : cellular morphogenesis
UPB1 protein	7706509			
Beta-ureidopropionase (EC 3.5.1.6) (Beta-alanine synthase) (N- carbamoyl-beta-alanine amidohydrolase) (BUP-1)	7706509	0003824 : catalytic activity; 0016787 : hydrolase activity; 0043167 : ion binding		0006807 : nitrogen compound metabolism
Alpha-1-acid glycoprotein 1 precursor (AGP 1) (Orosomucoid-1) (OMD 1)	9257232	0005488 : binding; 0005515 : protein binding	0005576 : extracellular region	0050896 : response to stimulus
Orosomucoid 1	9257232	0005488 : binding		

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Tuftelin	9910596	0005198 : structural molecule activity; 0005515 : protein binding	0005737 : cytoplasm; 0005576 : extracellular region	0007275 : development
ATP-binding cassette sub-family B member 10, mitochondrial precursor (ATP-binding cassette transporter 10) (ABC transporter 10 protein) (Mitochondrial ATP-binding cassette 2) (M-ABC2)	9961244	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity	0016020 : membrane; 0005739 : mitochondrion; 0043190 : ATP-binding cassette (ABC) transporter complex	0006810 : transport
Hemopexin precursor (Beta-1B-glycoprotein)	11321561	0005215 : transporter activity; 0005488 : binding; 0043167 : ion binding	0005576 : extracellular region	0019725 : cell homeostasis; 0051181 : cofactor transport; 0006810 : transport
Integral membrane protein 2C (Transmembrane protein BRI3) (Cerebral protein 14) [Contains: CT-BRI3]	13569885	0000166 : nucleotide binding	0016020 : membrane; 0005794 : Golgi apparatus	
Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona occludens 2 protein) (Tight junction protein 2)	13634076	0005515 : protein binding; 0016740 : transferase activity	0016020 : membrane; 0005634 : nucleus	
Allograft inflammatory factor 1 (AIF-1) (Ionized calcium-binding adapter molecule 1) (Protein G1)	14574568	0043167 : ion binding	0005634 : nucleus	0008283 : cell proliferation; 0050789 : regulation of biological process; 0050896 : response to stimulus; 0007049 : cell cycle
Allograft inflammatory factor 1 (Allograft inflammatory factor 1 isoform 3)	14574568	0043167 : ion binding		

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Cullin-5 (CUL-5) (Vasopressin-activated calcium-mobilizing receptor) (VACM-1)	14917099	0004871 : signal transducer activity; 0005515 : protein binding; 0005215 : transporter activity		0008283 : cell proliferation; 0050789 : regulation of biological process; 0007049 : cell cycle; 0006464 : protein modification; 0008219 : cell death
Keratin, type I cytoskeletal 14 (Cytokeratin-14) (CK-14) (Keratin-14) (K14)	15431310	0005198 : structural molecule activity; 0005515 : protein binding	0005856 : cytoskeleton	0007275 : development
Spectrin beta chain, brain 3 (Spectrin, non-erythroid beta chain 3) (Beta-IV spectrin)	17368942	0005515 : protein binding; 0005198 : structural molecule activity	0016020 : membrane; 0005737 : cytoplasm; 0005634 : nucleus; 0043005 : neuron projection; 0005856 : cytoskeleton; 0043025 : cell soma	0016043 : cell organization and biogenesis; 0050789 : regulation of biological process; 0051261 : protein depolymerization; 0007267 : cell-cell signaling; 0016192 : vesicle-mediated transport; 0008150 : biological_process
Monocarboxylate transporter 3 (MCT 3) (Solute carrier family 16 member 8)	17433295	0005215 : transporter activity	0016020 : membrane; 0000267 : cell fraction	0006810 : transport; 0015849 : organic acid transport; 0006811 : ion transport
ATP-binding cassette sub-family G member 1 (White protein homolog) (ATP-binding cassette transporter 8)	17433715	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005215 : transporter activity; 0005515 : protein binding	0016020 : membrane; 0005783 : endoplasmic reticulum; 0000267 : cell fraction	0006869 : lipid transport; 0006810 : transport; 0009987 : cellular process; 0050896 : response to stimulus; 0006066 : alcohol metabolism; 0006629 : lipid metabolism; 0042592 : homeostasis
ESF1 homolog (ABT1-associated protein)	18093112		0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
AP-3 complex subunit beta-2 (Adapter-related protein complex 3 beta-2 subunit) (Beta3B-adaptin) (Adaptor protein complex AP-3 beta-2 subunit) (AP-3 complex beta-2 subunit) (Clathrin assembly protein complex 3 beta-2 large chain) (Neuron-specific vesicle coat protein beta-NAP)	18202497	0005488 : binding; 0005215 : transporter activity	0016020 : membrane; 0005794 : Golgi apparatus; 0016023 : cytoplasmic membrane-bound vesicle	0015031 : protein transport; 0006810 : transport; 0016044 : membrane organization and biogenesis; 0016192 : vesicle-mediated transport; 0045045 : secretory pathway; 0046907 : intracellular transport
Enhancer of zeste homolog 1 (ENX-2)	19923202	0003676 : nucleic acid binding; 0005488 : binding	0005634 : nucleus	0008150 : biological_process; 0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Enhancer of zeste homolog 1 (Drosophila)	19923202	0003676 : nucleic acid binding	0005634 : nucleus	
HCTP4	20127519			
Targeting protein for Xklp2 (Restricted expression proliferation- associated protein 100) (p100) (Differentially expressed in cancerous and noncancerous lung cells 2) (DIL-2) (Protein fls353) (Hepatocellular carcinoma-associated antigen 519)	20127519	0000166 : nucleotide binding; 0005515 : protein binding	0005819 : spindle; 0005856 : cytoskeleton; 0005634 : nucleus	0007049 : cell cycle; 0008283 : cell proliferation
Prestin (Solute carrier family 26 member 5)	20139418	0003774 : motor activity; 0005215 : transporter activity	0016020 : membrane	0000902 : cellular morphogenesis; 0050789 : regulation of biological process; 0006810 : transport; 0006811 : ion transport; 0008150 : biological_process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
ATP-citrate synthase (EC 2.3.3.8) (ATP-citrate (pro-S-)-lyase) (Citrate cleavage enzyme)	20141248	0003824 : catalytic activity; 0043167 : ion binding; 0000166 : nucleotide binding; 0016740 : transferase activity	0005737 : cytoplasm; 0043234 : protein complex	0006629 : lipid metabolism; 0008152 : metabolism; 0005975 : carbohydrate metabolism; 0006082 : organic acid metabolism; 0051186 : cofactor metabolism; 0009117 : nucleotide metabolism
Protein NDRG2 (Protein Syld709613)	20141615		0005737 : cytoplasm	0007275 : development; 0009987 : cellular process
Lysosomal protective protein precursor (EC 3.4.16.5) (Cathepsin A) (Carboxypeptidase C) (Protective protein for beta-galactosidase) [Contains: Lysosomal protective protein 32 kDa chain; Lysosomal protective protein 20 kDa chain]	20178316	0016787 : hydrolase activity; 0005215 : transporter activity; 0005515 : protein binding; 0030234 : enzyme regulator activity	0005773 : vacuole; 0005739 : mitochondrion; 0005783 : endoplasmic reticulum	0015031 : protein transport; 0046907 : intracellular transport; 0019538 : protein metabolism
Cytoplasmic protein NCK2 (NCK adaptor protein 2) (SH2/SH3 adaptor protein NCK-beta) (Nck-2)	20532395	0005515 : protein binding	0005737 : cytoplasm	0008283 : cell proliferation; 0050789 : regulation of biological process; 0009987 : cellular process; 0007165 : signal transduction; 0006461 : protein complex assembly; 0016043 : cell organization and biogenesis; 0019538 : protein metabolism

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Homeodomain-interacting protein kinase 2 (EC 2.7.11.1) (hHIPk2)	21431782	0016740 : transferase activity; 0000166 : nucleotide binding; 0005515 : protein binding; 0030528 : transcription regulator activity; 0046790 : virion binding	0005737 : cytoplasm; 0005634 : nucleus	0007165 : signal transduction; 0050789 : regulation of biological process; 0008219 : cell death; 0006350 : transcription; 0016070 : RNA metabolism; 0008150 : biological_process; 0007049 : cell cycle; 0006464 : protein modification; 0006793 : phosphorus metabolism; 0050896 : response to stimulus
Protein S100-A8 (S100 calcium-binding protein A8) (Calgranulin-A) (Migration inhibitory factor-related protein 8) (MRP-8) (Cystic fibrosis antigen) (CFAG) (P8) (Leukocyte L1 complex light chain) (Calprotectin L1L subunit) (Urinary stone protein band A)	21614544	0043167 : ion binding	0005576 : extracellular region	0050896 : response to stimulus
Estrogen sulfotransferase (EC 2.8.2.4) (Sulfotransferase, estrogen- preferring) (EST-1)	21730331	0005496 : steroid binding; 0016740 : transferase activity; 0008289 : lipid binding		0006629 : lipid metabolism; 0042445 : hormone metabolism; 0008150 : biological_process
Serine/threonine-protein kinase Nek1 (EC 2.7.11.1) (NimA-related protein kinase 1) (Renal carcinoma antigen NY-REN-55)	22256934	0016740 : transferase activity; 0043167 : ion binding; 0000166 : nucleotide binding; 0005515 : protein binding	0005737 : cytoplasm; 0005634 : nucleus	0007049 : cell cycle; 0051301 : cell division; 0006464 : protein modification; 0006793 : phosphorus metabolism
Bromodomain adjacent to zinc finger domain protein 2B (hWALp4)	22653668	0003676 : nucleic acid binding; 0005515 : protein binding; 0043167 : ion binding	0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Lethal(3)malignant brain tumor-like protein (L(3)mbt-like) (L(3)mbt protein homolog) (H-l(3)mbt protein) (H-L(3)MBT) (L3MBTL1)	23396689	0003676 : nucleic acid binding; 0030528 : transcription regulator activity; 0043167 : ion binding	0005634 : nucleus	0006323 : DNA packaging; 0006350 : transcription; 0050789 : regulation of biological process; 0016070 : RNA metabolism
Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa ribosome receptor homolog) (ES/130-related protein)	23822112	0004871 : signal transducer activity; 0005198 : structural molecule activity	0016020 : membrane; 0005840 : ribosome; 0005783 : endoplasmic reticulum	0015031 : protein transport; 0046907 : intracellular transport; 0007165 : signal transduction; 0006412 : protein biosynthesis; 0006810 : transport
Dedicator of cytokinesis protein 9 (Cdc42 guanine nucleotide exchange factor zizimin-1)	24308029	0030234 : enzyme regulator activity; 0005515 : protein binding; 0000166 : nucleotide binding		
Glutathione S-transferase A3 (EC 2.5.1.18) (Glutathione S-transferase A3-3) (GST class-alpha member 3)	24430144	0016740 : transferase activity		0008152 : metabolism; 0050896 : response to stimulus
Glutathione S-transferase A3	24430144	0016740 : transferase activity		0008152 : metabolism
Fibrinogen beta chain precursor [Contains: Fibrinopeptide B]	24987624	0005515 : protein binding	0000267 : cell fraction; 0005576 : extracellular region; 0043234 : protein complex	0050817 : coagulation; 0050896 : response to stimulus; 0007165 : signal transduction; 0009987 : cellular process; 0019538 : protein metabolism; 0008150 : biological_process; 0008283 : cell proliferation; 0050789 : regulation of biological process
Keratin, type II cytoskeletal 6E (Cytokeratin-6E) (CK 6E) (K6e keratin) (Keratin K6h)	27465517	0005198 : structural molecule activity	0005856 : cytoskeleton	0016043 : cell organization and biogenesis

Appendix IX Cont'd

Protein Name	Matched Accession No	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Apolipoprotein A-I precursor (Apo-AI) (ApoA-I) [Contains: Apolipoprotein A-I(1-242)]	2914175; 4557321	0005215 : transporter activity; 0005515 : protein binding; 0008289 : lipid binding	0005576 : extracellular region	0006869 : lipid transport; 0019538 : protein metabolism; 0050789 : regulation of biological process; 0006629 : lipid metabolism; 0006810 : transport; 0008150 : biological_process; 0006066 : alcohol metabolism
Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal- binding globulin)	4389232; 4557871	0043167 : ion binding	0005576 : extracellular region; 0005768 : endosome; 0016023 : cytoplasmic membrane-bound vesicle	0006810 : transport; 0019725 : cell homeostasis; 0006811 : ion transport

APPENDIX X List of CSF *falciparum* proteins with Gene Ontologies

Protein Name	Matched Fields	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein PFB0755w	PFB0755w			
Hypothetical protein	PFL1715w	0016740 : transferase activity	0005634 : nucleus	0006139 : nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism
DHHC-type zinc finger protein, putative	PFI1580c	0043167 : ion binding		
Hypothetical protein PFB0770c	PFB0770c			
DNA repair protein, putative	PFE0270c	0003676 : nucleic acid binding; 0000166 : nucleotide binding		0006260 : DNA replication; 0006281 : DNA repair; 0006259 : DNA metabolism
Hypothetical protein	PFF0125c			
Hypothetical protein PFB0765w	PFB0765w			
Hypothetical protein	PFL1330c			
Hypothetical protein PFI0175w	PFI0175w			
Hypothetical protein	PFL2505c			
Hypothetical protein PFI1120c	PFI1120c			
Hypothetical protein PFE1095w	PFE1095w			
Hypothetical protein PFD0200c	PFD0200c			
DNA polymerase epsilon, catalytic subunit a, putative (Putative dna polymerase epsilon, catalytic subunit a) (EC 2.7.7.7)	PFF1470c	0016740 : transferase activity; 0003676 : nucleic acid binding; 0016787 : hydrolase activity; 0000166 : nucleotide binding		0006260 : DNA replication

Appendix X Cont'd

Protein Name	Matched Fields	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
GTP cyclohydrolase I	PFL1155w	0016787 : hydrolase activity	0005737 : cytoplasm	0006725 : aromatic compound metabolism
Hypothetical protein	PFL0410w			
Hypothetical protein PFI0260c	PFI0260c	0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0016787 : hydrolase activity; 0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process; 0046907 : intracellular transport
CutA, putative	PFL2375c			
Hypothetical protein PFE0240c (Fragment)	PFE0240c			
Putative SET-domain protein (Set-domain protein, putative)	PFF1440w	0005515 : protein binding; 0043167 : ion binding	0005634 : nucleus	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process
Hypothetical protein PFI0665w	PFI0665w			
Hypothetical protein PF13_0078	PF13_0078			
Hypothetical protein PF13_0091	PF13_0091			
Hypothetical protein PF08_0080	PF08_0080			
Hypothetical protein MAL8P1.103	MAL8P1.103			
Hypothetical protein MAL8P1.139	MAL8P1.139			
Hypothetical protein	PF11_0468			
Hypothetical protein	PF11_0374	0003676 : nucleic acid binding		
Hypothetical protein Phat31	PF13_0292			
Cg1 protein	PF07_0035			

Appendix X Cont'd

Protein Name	Matched Fields	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein MAL7P1.167	MAL7P1.167			
Hypothetical protein	PF11_0392			
Hypothetical protein PF07_0086	PF07_0086			
Phosphatidylinositol transfer protein, putative	MAL13P1.256		0005622 : intracellular	0006810 : transport
Queuine tRNA-ribosyltransferase; putative (EC 2.4.2.29)	PF07_0071	0016740 : transferase activity		0009451 : RNA modification; 0009116 : nucleoside metabolism
Hypothetical protein MAL13P1.234	MAL13P1.234			
Hypothetical protein MAL8P1.59	MAL8P1.59	0003676 : nucleic acid binding	0030529 : ribonucleoprotein complex	
Hypothetical protein	PF11_0129			
Hypothetical protein PF13_0027	PF13_0027			
Hypothetical protein PFD1160w	PFD1160w			
Hypothetical protein MAL13P1.144	MAL13P1.144	0016740 : transferase activity		
Hypothetical protein MAL7P1.134	MAL7P1.134			
Dynein beta chain, putative	PF14_0626	0003774 : motor activity	0005856 : cytoskeleton; 0043234 : protein complex	0046907 : intracellular transport
Hypothetical protein	PF11_0115			

Appendix X Cont'd

Protein Name	Matched Fields	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
Hypothetical protein PFD0535w	PFD0535w			
Hypothetical protein	PF14_0169	0003676 : nucleic acid binding	0005622 : intracellular	
Hypothetical protein	PF14_0419	0016787 : hydrolase activity; 0043167 : ion binding		0019538 : protein metabolism; 0009405 : pathogenesis
DNA repair helicase, putative	PF14_0081	0016787 : hydrolase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0003824 : catalytic activity		0006139 : nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism
Hypothetical protein	PF14_0175			
NAD(P)H-dependent glutamate synthase, putative	PF14_0334	0003824 : catalytic activity; 0043167 : ion binding; 0016491 : oxidoreductase activity	0016020 : membrane	0006091 : generation of precursor metabolites and energy; 0008152 : metabolism; 0006520 : amino acid metabolism; 0006807 : nitrogen compound metabolism
Hypothetical protein	PF14_0712			
Hypothetical protein	PF14_0320	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Ser/Thr protein kinase, putative	PF14_0392	0016740 : transferase activity; 0000166 : nucleotide binding		0006464 : protein modification; 0006793 : phosphorus metabolism
Hypothetical protein	PF14_0509			
QF122 antigen	PF10_0115	0003676 : nucleic acid binding		
Hypothetical protein	PF10_0267			
Hypothetical protein	PF14_0594			
Rifin	PF14_0002			
Cleavage and polyadenylation specifity factor protein, putative	PF14_0364	0016787 : hydrolase activity		
Hypothetical protein	PF14_0372			

Appendix X Cont'd

Protein Name	Matched Fields	GO Slim Func.	GO Slim Comp.	GO Slim Proc.
POM1, putative	PF14_0112	0016740 : transferase activity; 0003676 : nucleic acid binding; 0000166 : nucleotide binding; 0016787 : hydrolase activity; 0003824 : catalytic activity	0005622 : intracellular	0006260 : DNA replication; 0006304 : DNA modification
Hypothetical protein	PF14_0326	0000166 : nucleotide binding; 0016787 : hydrolase activity; 0005515 : protein binding	0005622 : intracellular	0006350 : transcription; 0016070 : RNA metabolism; 0050789 : regulation of biological process

APPENDIX XI Example of R version 2.5.1 script used to generate representative spectra in chapter 6

```
zap()
rm(.data)

## Spectrum for abm

use("abm.dta")

pdf(file="abm.pdf")

plot(.data, type="l", xlab="mass in daltons")
title("a)abm raw spectrum")

plot(.data, type="l", xlim=c(5000,50000),xlab="mass in daltons")
title("b)abm raw spectrum")

plot(.data, type="l", xlim=c(10000,20000),xlab="mass in daltons")
title("b)abm raw spectrum")

plot(.data, type="l", xlab="mass in daltons", xlim=c(64000,69000))
title("c)abm raw spectrum with restricted mass values")

bseoff<-bslnoff(.data, method="loess", bw = 0.1,xlim=c(7000,150000),
      xlab= "mass in daltons", plot = TRUE)
title(" c)Spectrum with baseline removed")

dev.off()
```